

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, C12N 1/21, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, 33/53, 33/68, A61K 38/17		A2	(11) International Publication Number: WO 98/39448 (43) International Publication Date: 11 September 1998 (11.09.98)																														
(21) International Application Number: PCT/US98/04493 (22) International Filing Date: 6 March 1998 (06.03.98) (30) Priority Data: <table border="0"><tr><td>60/040,162</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,333</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/038,621</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,161</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,626</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,334</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,336</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,163</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/043,580</td><td>11 April 1997 (11.04.97)</td><td>US</td></tr><tr><td>60/043,568</td><td>11 April 1997 (11.04.97)</td><td>US</td></tr></table> <p><i>(Continued on the following page)</i></p> (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hills Road, Laytonsville, MD 20882 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPER, Daniel, R. [US/US]; 15050 Stillfield, Place, Centreville, VA 22020 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). BEDNARIK, Daniel, P. [US/US]; 8822 Blue Sea Drive, Columbia, MD 21046 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). YOUNG, Paul, E. [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown,		60/040,162	7 March 1997 (07.03.97)	US	60/040,333	7 March 1997 (07.03.97)	US	60/038,621	7 March 1997 (07.03.97)	US	60/040,161	7 March 1997 (07.03.97)	US	60/040,626	7 March 1997 (07.03.97)	US	60/040,334	7 March 1997 (07.03.97)	US	60/040,336	7 March 1997 (07.03.97)	US	60/040,163	7 March 1997 (07.03.97)	US	60/043,580	11 April 1997 (11.04.97)	US	60/043,568	11 April 1997 (11.04.97)	US	MD 20874 (US). DUAN, Roxanne [US/US]; 4541 Fairfield Drive, Bethesda, MD 20814 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mount Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment #104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [BU/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). (74) Agents: BROOKES, Anders, A. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
60/040,162	7 March 1997 (07.03.97)	US																															
60/040,333	7 March 1997 (07.03.97)	US																															
60/038,621	7 March 1997 (07.03.97)	US																															
60/040,161	7 March 1997 (07.03.97)	US																															
60/040,626	7 March 1997 (07.03.97)	US																															
60/040,334	7 March 1997 (07.03.97)	US																															
60/040,336	7 March 1997 (07.03.97)	US																															
60/040,163	7 March 1997 (07.03.97)	US																															
60/043,580	11 April 1997 (11.04.97)	US																															
60/043,568	11 April 1997 (11.04.97)	US																															
(54) Title: 186 HUMAN SECRETED PROTEINS (57) Abstract <p>The present invention relates to 186 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.</p>		Published Without international search report and to be republished upon receipt of that report.																															

=> s wo9839448/pn

L1 1 W09839448/PN

=> d bib

Full
Text

OF 1 WPIDS COPYRIGHT 2004 THOMSON DERWENT on STN

AN 1998-506364 [43] WPIDS
CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35];
1998-427559 [36]; 1998-520811 [44]; 1998-609887 [51]; 1999-059865 [05];
1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11];
1999-190160 [16]; 1999-204988 [17]; 1999-418749 [35]; 1999-430031 [36];
1999-551363 [46]; 2000-106100 [09]; 2000-126931 [11]; 2000-161128 [14];
2000-182442 [16]; 2000-195282 [17]; 2000-482826 [42]; 2000-665238 [64];
2001-425865 [45]; 2001-625724 [72]; 2002-362489 [39]; 2002-574454 [61];
2002-598780 [64]; 2002-599716 [64]; 2002-634796 [68]; 2002-730795 [79];
2003-310989 [30]; 2003-466138 [44]; 2003-492322 [46]; 2003-511926 [48];
2003-521800 [49]; 2003-531736 [50]; 2003-540138 [51]; 2003-540785 [51];
2003-540804 [51]; 2003-567105 [53]; 2003-576674 [54]; 2003-829564 [77];
2003-864797 [80]; 2003-898535 [82]; 2003-901099 [82]; 2004-042167 [04];
2004-088563 [09]; 2004-131264 [13]; 2004-180094 [17]; 2004-225733 [21]
DNN N1998-394741 DNC C1998-152795
TI New isolated human genes and the secreted polypeptide(s) they encode -
useful for diagnosis and treatment of e.g. cancers, neurological
disorders, immune diseases, inflammation or blood disorders.
DC B04 D13 D16 S03
IN BEDNARIK, D P; BREWER, L A; CARTER, K C; DUAN, R; EBNER, R; ENDRESS, G A;
FENG, P; FERRIE, A M; FISCHER, C L; FLORENCE, K A; GREENE, J M; HU, J; NI,
J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SOPPET, D R; YOUNG, P E; YU, G;
DUAN, D R; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; SHI, Y; ZENG, Z;
DILLON, P J; FAN, P; FLORENCE, C; FLORENCE, K; WEI, Y; YOUNG, P
PA (HUMA-N) HUMAN GENOME SCI INC; (BREW-I) BREWER L A; (EBNE-I) EBNER R;
(FISC-I) FISCHER C L; (KYAW-I) KYAW H; (LAFL-I) LAFLEUR D W; (LIYY-I) LI
Y; (MOOR-I) MOORE P A; (OLSE-I) OLSEN H S; (ROSE-I) ROSEN C A; (RUBE-I)
RUBEN S M; (SHIY-I) SHI Y; (SOPP-I) SOPPET D R; (ZENG-I) ZENG Z
CYC 82
PI WO 9839448 A2 19980911 (199843)* EN 721<--
RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA
PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE
GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG
MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG
US UZ VN YU ZW
AU 9865453 A 19980922 (199908)
AU 9891304 A 19990322 (199931)
EP 972030 A2 20000119 (200009) EN
R: AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
US 6342581 B1 20020129 (200210)
JP 2002510192 W 20020402 (200225)
US 2003022185 A1 20030130 (200311)
US 2003092893 A1 20080515 (200335)
US 2003166906 A1 20030904 (200359)
EP 1352962 A1 20031015 (200368) EN
R: AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
ADT WO 9839448 A2 WO 1998-US4493 19980306; AU 9865453 A AU 1998-65453
19980306; AU 9891304 A AU 1998-91304 19980903; EP 972030 A2 EP 1998-905127
19980306, WO 1998-US4493 19980306; US 6342581 B1 Provisional US
1997-51916P 19970708, Provisional US 1997-51918P 19970708, Provisional US
1997-51919P 19970708, Provisional US 1997-51920P 19970708, Provisional US
1997-51925P 19970708, Provisional US 1997-51926P 19970708, Provisional US

60/043,314	11 april 1997 (11.04.97)	US	60/047,596	23 May 1997 (23.05.97)	US	60/056,637	22 August 1997 (22.08.97)	US
60/043,569	11 april 1997 (11.04.97)	US	60/047,612	23 May 1997 (23.05.97)	US	60/056,903	22 August 1997 (22.08.97)	US
60/043,311	11 april 1997 (11.04.97)	US	60/047,632	23 May 1997 (23.05.97)	US	60/056,888	22 August 1997 (22.08.97)	US
60/043,671	11 april 1997 (11.04.97)	US	60/047,601	23 May 1997 (23.05.97)	US	60/056,879	22 August 1997 (22.08.97)	US
60/043,674	11 april 1997 (11.04.97)	US	60/047,595	23 May 1997 (23.05.97)	US	60/056,880	22 August 1997 (22.08.97)	US
60/043,669	11 april 1997 (11.04.97)	US	60/047,599	23 May 1997 (23.05.97)	US	60/056,894	22 August 1997 (22.08.97)	US
60/043,312	11 april 1997 (11.04.97)	US	60/047,588	23 May 1997 (23.05.97)	US	60/056,911	22 August 1997 (22.08.97)	US
60/043,313	11 april 1997 (11.04.97)	US	60/047,585	23 May 1997 (23.05.97)	US	60/056,636	22 August 1997 (22.08.97)	US
60/043,672	11 april 1997 (11.04.97)	US	60/047,586	23 May 1997 (23.05.97)	US	60/056,874	22 August 1997 (22.08.97)	US
60/043,578	11 april 1997 (11.04.97)	US	60/047,590	23 May 1997 (23.05.97)	US	60/056,910	22 August 1997 (22.08.97)	US
60/043,576	11 april 1997 (11.04.97)	US	60/047,594	23 May 1997 (23.05.97)	US	60/056,864	22 August 1997 (22.08.97)	US
60/043,670	11 april 1997 (11.04.97)	US	60/047,589	23 May 1997 (23.05.97)	US	60/056,631	22 August 1997 (22.08.97)	US
60/047,600	23 May 1997 (23.05.97)	US	60/047,593	23 May 1997 (23.05.97)	US	60/056,845	22 August 1997 (22.08.97)	US
60/047,615	23 May 1997 (23.05.97)	US	60/047,614	23 May 1997 (23.05.97)	US	60/056,892	22 August 1997 (22.08.97)	US
60/047,597	23 May 1997 (23.05.97)	US	60/047,501	23 May 1997 (23.05.97)	US	60/056,632	22 August 1997 (22.08.97)	US
60/047,502	23 May 1997 (23.05.97)	US	60/048,974	06 June 1997 (06.06.97)	US	60/056,664	22 August 1997 (22.08.97)	US
60/047,633	23 May 1997 (23.05.97)	US	60/048,964	06 June 1997 (06.06.97)	US	60/056,876	22 August 1997 (22.08.97)	US
60/047,583	23 May 1997 (23.05.97)	US	60/049,610	13 June 1997 (13.06.97)	US	60/056,881	22 August 1997 (22.08.97)	US
60/047,617	23 May 1997 (23.05.97)	US	60/051,926	08 July 1997 (08.07.97)	US	60/056,909	22 August 1997 (22.08.97)	US
60/047,618	23 May 1997 (23.05.97)	US	60/052,874	16 July 1997 (16.07.97)	US	60/056,875	22 August 1997 (22.08.97)	US
60/047,503	23 May 1997 (23.05.97)	US	60/055,724	18 August 1997 (18.08.97)	US	60/056,862	22 August 1997 (22.08.97)	US
60/047,592	23 May 1997 (23.05.97)	US	60/056,886	22 August 1997 (22.08.97)	US	60/056,887	22 August 1997 (22.08.97)	US
60/047,581	23 May 1997 (23.05.97)	US	60/056,877	22 August 1997 (22.08.97)	US	60/056,908	22 August 1997 (22.08.97)	US
60/047,584	23 May 1997 (23.05.97)	US	60/056,889	22 August 1997 (22.08.97)	US	60/056,884	22 August 1997 (22.08.97)	US
60/047,500	23 May 1997 (23.05.97)	US	60/056,893	22 August 1997 (22.08.97)	US	60/057,761	05 September 1997 (05.09.97)	US
60/047,587	23 May 1997 (23.05.97)	US	60/056,630	22 August 1997 (22.08.97)	US	60/057,650	05 September 1997 (05.09.97)	US
60/047,492	23 May 1997 (23.05.97)	US	60/056,878	22 August 1997 (22.08.97)	US	60/057,669	05 September 1997 (05.09.97)	US
60/047,598	23 May 1997 (23.05.97)	US	60/056,662	22 August 1997 (22.08.97)	US	60/058,785	12 September 1997 (12.09.97)	US
60/047,613	23 May 1997 (23.05.97)	US	60/056,872	22 August 1997 (22.08.97)	US	60/061,060	02 October 1997 (02.10.97)	US
60/047,582	23 May 1997 (23.05.97)	US	60/056,882	22 August 1997 (22.08.97)	US			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

186 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table I, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 12301 Park Lawn Drive, Rockville, Maryland 20852, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA contained within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency), salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6x SSPE (20x SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1xSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5x SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, uridyated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA, thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslational natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

25 Polynucleotides and Polypeptides of the Invention

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in testes tumor and to a lesser extent in fetal brain. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly of the testes, and defects of the central nervous system such as seizure and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly cancer of the testes and central nervous system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, brain and other tissue of the nervous system, and blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of testicular cancer and treatment of central nervous system disorders since this gene is primarily expressed in the testes tumor and developing brain.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in cancer tissues, such as breast cancer and Wilms' tumor, and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and/or tumors, particularly, those found in the breast, and developmental abnormalities or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and fetal tissue and, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 314 as residues: Pro-11 to Thr-18, Leu-43 to Pro-50, Gly-64 to Leu-72, and Leu-81 to Lys-86.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of cancers and/or tumors, particularly, those found in the breast since expression is mainly in cancer/tumor tissues. May serve as therapeutic proteins for proliferation/differentiation of fetal tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in spleen, chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or leukemias, diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders or leukemias, diseases of the immune system since expression is in tissues related to immune function.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or lymphocytic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 317 as residues: Pro-13 to Lys-21.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 6

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 318 as residues: Lys-31 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood diseases since it is expressed in tissues related to immune function.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 7

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system and brain associated diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders, immune diseases or brain associated diseases (specifically of the pineal gland) since expression is in tissues related to immune function.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 8

The translation product of this gene shares sequence homology with an organic cation transporter which is thought to be important in organic cation uptake in the kidney and liver. (See Accession No. 2343059.) Preferred polypeptide fragments comprise the amino acid sequence ITAIQMICLVNXXEL YPTFVRNXGVNVCSLDCIGIITP FIVFRLREVWQALPLILFAVLGLLAAGVTLLPETKGVLPETMKDAENLGRKAKPKENTLYLK VQTSEPSQT (SEQ ID NO: 615) or TMKDAENLGRKAKPKENT (SEQ ID NO: 616) as well as N-terminal and C-terminal deletions of these fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic and renal diseases where drug elimination/cation exchange (organic cation uptake) in the liver and kidney are problematic. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 320 as residues: Asn-64 to Asn-74, and Gln-81 to Gly-87.

The tissue distribution and homology to organic cation transporter indicate that polynucleotides and polypeptides corresponding to this gene are useful as a polyspecific transporter that is important for drug elimination in the liver (and possibly kidney) since expression is found in the liver.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in eosinophil induced with IL-5 and to a lesser extent in fetal liver and spleen. This gene also maps to chromosome 15, and therefore can be used in linkage analysis as a marker for chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system, particularly allergies or asthma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosis of diseases involving eosinophil reactions since expression seems to be concentrated in eosinophils and other tissues involved in immunity such as the liver and spleen.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

This gene is expressed primarily in tissues of hematopoietic lineage and to a lesser extent in Hodgkins lymphoma. Any frame shifts in this sequence can easily be clarified using known molecular biology techniques.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and immune deficiency or dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, lymphoid and reticuloendothelial tissues, and cancerous tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/ diagnosis for lymphomas or immune dysfunction or as a therapeutic protein useful in immune modulation based on expression in anergic T-cells and lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in neutrophils and to a lesser extent in activated lymphoid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the cell type present in a biological sample and for diagnosis of diseases and conditions: inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 323 as residues: Glu-40 to Lys-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for modulation of an immune reaction or as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in brain and to a lesser extent in activated T-cells. It is likely that the open reading frame containing the predicted signal peptide continues in the 5' direction. Preferred polypeptide fragments comprise the amino acid sequence PRVRNSPEDGLSLTGDSCKL (SEQ ID NO:617).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disorders including ischemic shock, alzheimers and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain, and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 324 as residues: Ser-5 to Glu-14, Ile-21 to Pro-35, Ser-65 to Asp-81, Cys-89 to Val-96, Lys-136 to Ser-145, Ile-152 to Met-169, and Arg-189 to Lys-196.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic/treatment for cancers of the given tissue or in the treatment of neurological disorders of the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene was also recently cloned by other groups, naming this calcium-activated potassium channel gene, hKCa4. (See Accession No. AF033021, see also, Accession No. 2584866.) This gene is mapped to human chromosome 19q13.2. A second signal sequence likely exists upstream from the predicted signal sequence as described in Table 1. Preferred polypeptide fragments comprise: QADDLQATVAALCYLRGGGPWAG SWLSPKTPGAMGGDLVLGLGALRRKRLL (SEQ NO: 618); or EQEKSLAGWALVLAXXGIOL MVLHAEMLWFGCSAVNATGHLSDTLWLIPITFLTIGYGDVVVPGTMWQKVCVLCCTGVMGVCC TALLVAVVARKLEFNKAEKHVHFMMDIQYTKEMKESAAARVLQEAWMFPYKHTRRKESHAAR XHQKXLLAAINAFRQVRLKRLREQVNSMVDISKMHMILYDLQNLSSSHRALEKQIDTL KLDALTELLSTALGPRQLPEPSQSK (SEQ ID NO: 619), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in breast lymph node and T-cells, and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematologic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, blood cells and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 325 as residues: Arg-13 to Lys-23.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment/diagnosis of hematologic and diseases involving immune modulation based on distribution in the lymph node and T-cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene was recently cloned by another group, calling it PAPS synthase.

(See Accession No. e1204135.) Preferred polypeptide fragments comprise the amino acid sequence YQAHFVSRNKRQGVVGTGRGFRCTVWLTLGSGAGK (SEQ ID NO: 620).

Also preferred are the polynucleotide fragments encoding this polypeptide fragment.

It has been discovered that this gene is expressed primarily in benign prostate hyperplasia, Human Umbilical Vein Endothelial Cells and to a lesser extent in smooth muscle and Human endometrial stromal cells-treated with estradiol.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammation, ischemia, and restenosis, based on endothelial cell and smooth muscle cell expression, and prostate diseases such as benign prostate hyperplasia or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s).

For a number of disorders of the above tissues or cells, particularly of the prostate or vessels of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, endothelial cells, smooth muscle, and endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 326 as residues: Arg-21 to Asp-26, Lys-35 to Lys-44, Glu-49 to Asn-58.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosing diseases or conditions where the endothelial cell lining of the veins and arteries of underlying smooth muscle are involved.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

This gene is expressed primarily in human 6 week embryo and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental anomalies or fetal deficiencies. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly developmental in nature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 327 as residues Lys-50 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of developmental abnormalities.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

This gene is expressed primarily in kidney and amygdala and to a lesser extent in fetal tissues. This gene is mapped to chromosome 14, and therefore is useful in linkage analysis as a marker for chromosome 14.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) present in a biological sample and for diagnosis of diseases and conditions: kidney diseases, neurological disorders and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s). For a number of disorders of the above tissues, particularly of the renal system or developing fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, amygdala, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of conditions affecting the brain, kidneys and fetal development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: solid tumors similar to ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 329 as residues Ser-51 to Val-56.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of solid tumors of the reproductive system such as ovarian cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in brain medulloblastoma. Preferred polypeptide fragments comprise the amino acid sequence: IRHEQHFNFSLEMHSGSLLFLPLQL ILLPVCAHLHEELNC (SEQ ID NO: 643) and SPTSEEKGHLLQAEHRHPWVAGALVGYSGLTLTTCGPTKPNYFLKLLQEMHIRAN (SEQ ID NO: 644), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the CNS or. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating medulloblastoma or similar tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed primarily in adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose tissues expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating obesity by regulating the function and number of adipocytes

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in B cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of the immune system with an emphasis on B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumors of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of B cell derived tumors based on its expression in b cell lymphomas

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in immune cells and to a lesser extent in fetal tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cells of the immune system, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:333 as residues Asp-10 to Pro-19, Ser-74 to Tyr-79, Glu-95 to Lys-110.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment of diseases involving alterations in T cell activity.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

It has been discovered that this gene is expressed primarily in ovarian tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumors of the reproductive organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian

and other reproductive tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 334 as residues: Leu-22 to Gln-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovarian tumors as it has only been identified in ovarian tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

It has been discovered that this gene is expressed primarily in fetal tissues and to a lesser extent in osteoclastoma cell line

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteoporosis or arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment of conditions of abnormal bone remodeling due to enhanced activity of osteoclasts. This may be useful as a specific marker for malignancies derived from osteoclasts or their precursors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with a periplasmic ribonuclease which is thought to be important in degrading extracellular polynucleotides

It has been discovered that this gene is expressed primarily in serum treated smooth muscle cells

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular disease such as stenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Gln-30 to Lys-36, and Pro-41 to Arg-48.

The tissue distribution and homology to ribonucleases indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of pathological conditions of smooth muscle associated with bacterial or viral infiltration

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

This gene is expressed primarily in Early Stage Human Brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain development and related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain development and related diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases affecting human brain development and related diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

It has been discovered that this gene is expressed primarily in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases and other diseases related to brain diseases, which may be caused by brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain diseases and other diseases related.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

It has been discovered that this gene is expressed primarily in Anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases, inflammatory diseases and diseases related to T lymph cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune diseases, inflammatory diseases and diseases related to T lymph cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for immune diseases, and inflammatory diseases and diseases related to T lymph cells.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with *Shigella flexneri* positive transcriptional regulator CtrR (ctrR) gene which is thought to be important in regulation of gene expression.

10

This gene is expressed primarily in human synovial sarcoma and normal human brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases particularly sarcomas of the synovium. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain and synovium and other related human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., synovial tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

25

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human synovial sarcoma and other related human brain diseases.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed in bone marrow, infant brain, fetal liver and spleen, prostate and to a lesser extent in pineal gland, adipose tissue, kidney, adrenal gland, umbilical vein endothelial cells, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases related to bone marrow or

35

hematoplasic tissues, prostate, kidney, adrenal gland, and cardiovascular tissue or organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases related to hematoplasic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, hematopoietic cells, pineal gland, adipose tissue, kidney, adrenal gland, endothelial cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5

10

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases related to hematoplasic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 30

This gene is expressed primarily in meningea and to a lesser extent in breast and adult brain.

20

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases of the meningea and related brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the meningea and related brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., miningea, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

30

35

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the meningea and related brain diseases.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed in meningea, fetal spleen, osteoblast and to a lesser extent in activated T-cells, endometrial stromal cells, fetal lung, HL-60, thymus, testis and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: meningeal disease, osteoporosis, immune diseases, and hematoplastic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the

15 above tissues or cells, particularly of the meningeal diseases, osteoporosis, immune diseases, and hematoplastic diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endometrium, lung, thymus, testis, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of meningeal, osteoporosis, immune diseases, hematoplastic diseases, testis diseases and lung diseases.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in human thymus and to a much lesser extent in infant brain, T-cells, smooth muscle, endothelial cells, bone marrow, human ovarian tumor and keratinocytes testes, osteoclastoma, breast, and tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases involving the thymus, particularly thymic cancer and diseases involving T-cell maturation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the thymus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, brain, and other tissue of the nervous system, blood cells, bone marrow, ovaries, and testes, and other reproductive tissue, mammary tissue, tonsils, melanocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thymus particularly thymic cancer and diseases involving T-cell maturation.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human tonsils, and placenta, and to a lesser extent in adipocytes, melanocyte, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases, immune diseases, and obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory diseases, immune diseases, and obesity, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tonsils, placenta, adipocytes, melanocytes, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases such as inflammation, immune diseases, and obesity.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene is expressed in activated T cells, and to a lesser extent in pituitary, testis, and breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases relating to T cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary, testes and other reproductive tissue, mammary tissue, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of immune disorders.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain, and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene is expressed primarily in infant brain.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in human ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the ovarian disorders such as those involving germ cells, ovarian follicles, stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovaropathy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene is expressed primarily in lymph node breast cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for used as a diagnostic marker for breast cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

This gene is expressed primarily in brain and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal disorders such as trauma, brain degeneration, and brain tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

This gene is expressed in early stage human embryo, adrenal gland tumor, and immune tissues such as fetal liver, fetal spleen, T-cell, and myeloid progenitor cell line and to a lesser extent in ovary, colon cancer, and a few other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis including adrenal gland tumor, colon cancer and various other tumors, developmental and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, early stage human tissues, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, blood cells, bone marrow, ovary and other reproductive tissue, and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of immune and developmental disorders, and tumorigenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene is expressed primarily in fetal lung, endothelial cells, liver, thymus and a few other immune tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders such as immune deficiency and autoimmune diseases, pulmonary diseases, liver diseases, and tumor metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal lung, liver, endothelial cells, and immune tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain

5 tissues and cell types (e.g., lung, endothelial cells, liver, thymus, and other tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune disorders and pulmonary and hepatic diseases. Its promoter may also be used for immune system and lung-specific gene therapies. The expression of this gene in endothelial cells indicates that it may also involve in angiogenesis which therefore may play role in tumor metastasis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

15 This gene is expressed primarily in liver, thyroid, parathyroid and to a lesser extent in fetal lung, stomach and early embryos.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic regulation, obesity, hepatic failure, hepatocellular tumors or thyroiditis and thyroid tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive/endocrine system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, thyroid, parathyroid, lung, stomach, and embryonic tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and the extracellular locations indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of digestive/endocrine disorders, including metabolic regulation, hepatic failure, malabsorption, gastritis and neoplasms.

FEATURES OF PROTEIN ENCODED BY GENE NO: 43

This gene is expressed primarily in Schizophrenic adult brain, pituitary, front cortex, hypothalamus and to a lesser extent in retina, adipose and stomach cancer and placenta.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal tissue, adipose, stomach, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nerve system, including schizophrenia, neurodegeneration, and neoplasia. Additionally, a secreted protein in brain may serve as an endocrine.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

25 The translation product of this gene shares sequence homology with GTP binding proteins which are thought to be important in signal transduction and protein transport.

This gene is expressed primarily in umbilical vein and microvascular endothelial cells, GM-CSF treated macrophage, anergic T cells, osteoblast, osteoclast, CD34+ cells and to a lesser extent in gall bladder.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: bone formation and growth, osteonecrosis, osteoporosis, angiogenesis and/or hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoiesis systems, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues and cell types (e.g., endothelial cells, blood cells, bone, and gall bladder, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of bone formation and growth, osteonecrosis, osteoporosis, and/or hematopoiesis because its involvement in the growth signaling or angiogenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with signal sequence receptor gamma subunit which is thought to be important in protein translocation on endoplasmic reticulum.

This gene is expressed primarily in adrenal gland, salivary gland, prostate, and to a lesser extent in endothelial cells and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: protein secretion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the secretory organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, salivary gland, prostate, endothelial cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to SSR gamma subunit indicate that polynucleotides and polypeptides corresponding to this gene are useful for endocrine disorders, prostate cancer, xerostomia or sialorrhea.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed primarily in osteoclastoma cells and to a lesser extent in melanocyte, amygdala, brain, and stomach.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ossification, osteoporosis, fracture, osteonecrosis, osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, amygdala, brain and other tissue of the nervous system, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in intervention of ossification, osteoporosis, fracture, osteonecrosis and osteosarcoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with proline rich proteins which is thought to be important in protein-protein interaction.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological and psychological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful in intervention

and detection of neurological diseases, including trauma, neoplasia, degenerative or metabolic conditions in the central nerve system. Additionally, the gene product may be secreted by the brain as an endocrine.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with the AOCB gene from *Aspergillus nidulans* which is important in asexual development.

This gene is expressed primarily in infant brain and to a lesser extent in the

developing embryo, trachea tumors, B-cell lymphoma and synovial sarcoma.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative diseases, leukemia and sarcoma's. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
15 tissues or cells, particularly of the brain and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, blood cells, trachea, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution in infant brain and sarcoma's and homology to a gene involved in a key step of eukaryotic development (fungal spore formation) indicates that the protein product of this clone could play a role in neurological diseases such as schizophrenia, particularly in infants. The existence of the gene in a B-cell lymphoma indicates the gene may be used in the treatment and detection of leukemia.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary disorders including lung cancer. Similarly, polypeptides and antibodies directed to these

35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution of this gene only in fetal lung indicates that it plays a key role in development of the pulmonary system. This would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung. It may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in hematopoietic cell types and fetal cells and to a lesser extent in all tissue types.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in the immune system and hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution of this gene predominantly in hematopoietic cells and in the developing embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of lymphomas and disease states affecting the immune system or hematopoiesis disorders such as leukemia, AIDS, arthritis and asthma.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene is expressed primarily in prostate and to a lesser extent in fetal spleen, fetal liver, infant brain and T cell leukemias.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate disorders, prostate cancer, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, and/or prostate gland expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, spleen, liver, brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in prostate indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection or treatment of prostate disorders or prostate cancer. Its distribution in fetal liver and fetal spleen indicates it may play a role in the immune system and its misregulation could lead to immune disorders such as leukemia, arthritis and asthma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

The translation product of this gene shares sequence homology with dynein. This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuro-degenerative diseases of the brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly neuro-degenerative diseases expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution in the brain and homology to dynein, a microtubule motor protein involved in the positioning of cellular organelles and molecules indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection/treatment of neurodegenerative diseases, such as Alzheimers, Huntingtons, Parkinsons diseases and shizophrenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with ubiquitin-conjugation protein, an enzyme which is thought to be important in the processing of the Huntingtons Disease causing gene.

This gene is expressed primarily in brain and to a lesser extent in activated macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disease states including Huntington's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of brain tissues. For a number of disorders of the above tissues or cells, particularly of the neurological systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution of this gene in the brain and its homology to a Huntington interacting protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the regulation of the expression of the Huntington disease gene and other neurodegenerative diseases including spinocerebellar ataxia types I and III, dentatorubropallidoluyisian and spinal bulbar muscular atrophy. In addition, the existence of elevated levels of free ubiquitin pools in Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis indicates that the ubiquitin pathway of protein degradation plays a role in these disease states. Thus, considering the gene described here is homologous to a ubiquitin-conjugation protein it may play a general role in neurodegenerative conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

This gene is expressed primarily in T-cells (anergic T-cells, resting T-Cells, apoptotic T-cells) and lymph node (breast), as well as brain (hypothalamus, hippocampus, pituitary, infant brain, early-stage brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune (e.g. immunodeficiencies, autoimmunities, inflammation, leukemias & lymphomas) and neurological (e.g. Alzheimer's disease, dementia, schizophrenia) disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous, hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, lymphoid tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention or detection of pathologies associated with the hematopoietic and immune systems, such as anemias (leukemias). In addition, the expression in brain (including fetal) might suggest a role in developmental brain defects, neuro-degenerative diseases or behavioral abnormalities (e.g. schizophrenia, Alzheimer's, dementia, depression, etc.).

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in lung, and to a lesser extent in a variety of other hematological cell types (e.g. Raji cells, bone marrow cell line, activated monocytes).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary and/or hematological dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculo-pulmonary and hematopoietic systems, expression of this

gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention and detection of pathologies associated with the vasculo-pulmonary system. In addition the expression of this gene in a variety of leukocytic cell types and a bone marrow cell line might suggest a role in hematopoietic and immune system disorders, such as leukemias & lymphomas, inflammation, immunodeficiencies and autoimmunities.

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

The translation product of this gene shares sequence homology with adenylate kinase isozyme 3 (gill 63528 GTP:AMP phosphotransferase (EC 2.7.4.10) [Bos taurus]), which is thought to be important in catalyzing the phosphorylation of AMP to ADP in the presence of ATP or inorganic triphosphate.

This gene is expressed primarily in fetal liver, heart and placenta, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic, cardiovascular or reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic, cardiovascular and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, heart, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions related to hepatic function and pathogenesis, in particular, those dealing with liver development and the differentiation of hepatocyte progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in CD34 positive cells (Cord Blood).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic differentiation and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the detection and treatment of conditions associated with CD34-positive cells, and therefore as a marker for cell differentiation in hematopoiesis, as well as immunological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

The translation product of the predicted open reading frame of this contig has sequence identity to the murine gene designated Insulin-Like Growth Factor-Binding Protein (IGFBP)-1 as described by Lee and colleagues (Hepatology 19 (3), 656-665 (1994)).

This gene is expressed exclusively in hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hemangiopericytoma and other pericyte or endothelial cell proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and immune systems, expression of this gene at significantly higher or lower levels may routinely be detected in certain tissues and cell types (e.g., pericyte or endothelial cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Polynucleotides and polypeptides corresponding to this gene are useful as cell growth regulators since IGFBP-1-like molecules function as modulators of insulin-like growth factor activity. In addition, since IGFBP-1 is expressed at high levels following hepatectomy and during fetal liver development, polynucleotides of the present invention may also be used for the diagnosis of developmental disorders. Further, polypeptides of the present invention may be used therapeutically to treat developmental liver disorders as well as to regulate hepatocyte and supporting cell growth following hepatectomy or to treat liver disorders.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hemangiopericytoma and liver disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

This gene is expressed primarily in schizophrenic frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: nervous system and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the frontal cortex and CNS expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of frontal cortex, neuro-degenerative and CNS disorders

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in human adrenal gland tumor, and to a lesser extent in human kidney, medulla and adult pulmonary tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic, endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and nervous system disorders and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, kidney, brain and other tissue of the nervous system, pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of neurological and endocrine disorders including neoplasia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

This gene is expressed primarily in human adipocytes, and to a lesser extent in spleen, 12-week old human, and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune, metabolic and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes, spleen, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of immune, developmental and metabolic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 64

One translated product of this clone is homologous to the mouse zinc finger protein PZF. (See Accession No. 453376; see also Gene 152 (2), 233-238 (1995).) Preferred

polypeptide fragments correspond to the highly conserved domains shared between mouse and man. For example, preferred polypeptide fragments comprise the amino acid sequence: LQCEICGFTCRQKASLNWHMKKHADSFYQFSCNCGKKFEEKKDSVVVAKAKSH PEV (SEQ ID NO: 621); ITSDILGTNPESLTQPSD (SEQ ID NO: 622); NSTSGECLLLEAGQM SKSY (SEQ ID NO: 623); CSGTERVSLMADGKIFVGGSSGGTGLVMNSDILGATTEVLIEDSD SAGP (SEQ ID NO: 624); IQYVRCEMEGCGTVLAHPRYLQHHLKKKYVCPHPSCGRLP RLQQLLRHAKHHT (SEQ ID NO: 625); DQRDYCEYCARAFKSSHNLAVHRMIHTGEK (SEQ ID NO: 626); RSSRTSVSRHRDTESTRSSKSTGSLQICKSEPTDQLDY (SEQ ID NO: 627); PFDDPRDETYPHLERETPKRRKSG (SEQ ID NO: 630); QYVRCEMEGCGTVLAHPRYLQ HHIKYQHLLKKKYVCPHPSCORLFRLLQQLLRHAKHHTD (SEQ ID NO: 629); or residues 151-182 of QRDYCEYCARAFKSSHNLAVHRMIHTGEKHY (SEQ ID NO: 628). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Rhabdomyosarcoma, melanocyte and colon cancer tissue and to a lesser extent in smooth muscle, pancreatic tumor, and apoptotic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hemopoietic, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striated muscle, melanocytes, colon, smooth muscle, pancreas, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of cancer and hemopoietic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 65

This gene is expressed primarily in human adipose and salivary gland tissue and to a lesser extent in human bone marrow and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose, salivary gland, bone marrow, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for study, diagnosis of metabolic and immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

This translated product of this gene was recently identified as oxytocinase splice variant 1. (See Accession Nos. 2209276 and d1010078.) Preferred polypeptide fragments

comprise the amino acid sequence: EMFDLSYFKGSSLLMLKTYLSEDFVQHA VVLYLHN HSYASIQSDDLWDSFNEVTNQTLDYKRMKMTWTLQKGFPLVTVQKKGKELFQQRERFLNMK PEIQPSDTRYM (SEQ ID NO: 631). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene is expressed primarily in hemopoietic cells, particularly apoptotic T-cells, and to lesser extent in primary dendritic cells and adipose tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of apoptotic T-cells, primary dendritic cells, and adipose tissue present in a biological sample and for diagnosis of diseases and conditions: hemopoietic diseases including cancer and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the oral and intestinal mucosa as well as hemopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of diseases of the immune system, including cancer, hemopoietic and infectious diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 68

This gene is expressed primarily in kidney cortex and to a lesser extent in infant brain, heart, uterus, and blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of kidney tissue present in a biological sample and for diagnosis of diseases and conditions: soft tissue cancer, inflammation, kidney fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrines systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, brain, and other nervous tissue, heart, uterus, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and fibroses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

The translation product of this gene shares strong sequence homology with vertebrate and invertebrate protein tyrosine phosphatases.

This gene is expressed primarily in endometrial tumors, melanocytes, myeloid progenitors and to a lesser extent in infant brain, adipocytes, and several hematopoietic stem cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of transformed hematopoietic and epithelial cells present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of skin and endometrium, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, melanocytes, bone marrow, adipocytes, hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and sequence similarity with tyrosine phosphatases

indicate that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and hematopoietic disorders.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 70

This gene is expressed primarily in osteoclastoma, breast, and infant brain and to a lesser extent in various fetal and transformed bone, ovarian, and neuronal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: degenerative conditions of the brain and skeleton. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of degenerative, neurological and skeletal disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene was originally cloned from tumor cell lines. Recently another group has also cloned this gene, calling it the human malignant melanoma metastasis-suppressor (KISS-1) gene. (See Accession No. U43527.) Preferred polypeptide fragments comprise the amino acid sequence: LEKVASVGNRPQTQQLESIGLLA (SEQ ID NO: 632); VHREASCYCQAEPSGDL (SEQ ID NO: 633); RPALRQAGGOTREPRQKRWAGL (SEQ ID NO: 634); and AVNFRPQRSQSM (SEQ ID NO: 635). Any frame shifts can easily be resolved using known molecular biology techniques.

This gene is expressed primarily in many types of carcinomas and to a lesser extent in many normal organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanomas, and other hyperproliferative disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of transformed organ tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. As a tumor suppressor gene, increase amounts of the polypeptide can be used to treat patients having a particular cancer.

The tissue distribution indicates that this gene and the translated product is useful for diagnosing and study of cancer.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene is expressed primarily in striatum and to a lesser extent in adipocytes and hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of striatal cells present in a biological sample and for diagnosis of diseases and conditions: neurological, fat and lysosomal storage

diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striatal tissue, adipocytes, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of neurodegenerative and growth disorders.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

This gene is expressed primarily in bone marrow stromal cells and to a lesser extent in smooth muscle, testes, endothelium, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of bone marrow present in a biological sample and for diagnosis of diseases and conditions: connective tissue and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, stromal cells, smooth muscle, testes and other reproductive tissue, endothelium, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of connective tissue and blood diseases.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed primarily in brain, fetal liver and lung and to a lesser extent in retina, spinal chord, activated T-cells and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of brain and regenerating liver present in a biological sample and for diagnosis of diseases and conditions: CNS and spinal chord injuries, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, pulmonary tissue, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of hematopoietic and neurological conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

The translation product of this gene shares sequence homology with GTP binding proteins (intracellular).

This gene is expressed primarily in bone marrow, brain, and melanocytes and to a lesser extent in various endocrine and hematopoietic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic and nervous system conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, melanocytes, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder.

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to nucleotide binding factors indicate that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of brain degenerative, skin and blood diseases.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in activated T-cells and to a lesser extent in retina, brain, and fetal bone.

10

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of activated T-cells and developing brain present in a biological sample and for diagnosis of diseases and conditions: immune deficiencies and skeletal and neuronal growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and skeletomuscular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, retinal tissue, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis, study and treatment of cancer, urogenital, and brain degenerative diseases.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in fetal liver, activated monocytes, osteoblasts and to a lesser extent in synovial, brain, and lymphoid tissues.

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of myeloid and lymphoid present in a biological sample and for diagnosis of diseases and conditions: inflammation, immune deficiencies, cancer. Similarly, polypeptides and antibodies directed to these

35

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skeleton, expression of this gene at significantly

higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, blood cells, bone, synovial tissue, brain and other tissue of the nervous system, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for study, diagnosis, and treatment of lymphoid and mesenchymal cancers and nervous system diseases.

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 78

The translation product of this gene shares sequence homology with polymerase polypeptide precursor which is thought to be important in DNA repair and replication

This gene is expressed primarily in infant brain and to a lesser extent in tumors and tumor cell lines

15

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, especially of the neural system and developing organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

25

30

The tissue distribution and homology to polymerase polypeptide precursor indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers especially of the neural system and developing organs

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed primarily in muscle and endothelial cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., muscle, endothelial cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment and diagnosis of disorders of the vascular and neural system including cardiovascular and endothelial.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

This gene is expressed primarily in placenta and to a lesser extent in fetal liver

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental disorders and disorder of the haemopoietic system, fetal liver and placenta. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developmental disorders and disorder of the haemopoietic system, fetal liver and placenta, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of developmental disorders and disorders of the haemopoietic system, fetal liver and placenta.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed primarily in bone marrow, placenta and tissues and organs of the hematopoietic system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the bone and haemopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, bone and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, placenta, and hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of disorders of the immune, bone and hematopoietic system

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

The translation product of this gene shares sequence homology with secretory carrier membrane protein which is thought to be important in protein transport and export. Any frame shifts in coding sequence can be easily resolved using standard molecular biology techniques. Another group recently cloned this gene, calling it SCAMP. (See Accession No. 2232243.)

This gene is expressed primarily in prostate, breast and spleen, and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the breast prostate and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly disorders of the breast prostate and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., prostate, mammary tissue, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secretory carrier membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the breast, prostate and spleen.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed primarily in developing organs and tissue like placenta and infant brain and to a lesser extent in developed organs and tissue like cerebellum and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as

15 reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: neurological diseases.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, heart, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the neural system including neurological disorders and cancer.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

The translation product of this gene shares sequence homology with ATPase 6 in *Trypanosoma brucei* which is thought to be important in metabolism.

This gene is expressed primarily in tumor and fetal tissues and to a lesser extent in melanocytes, kidney cortex, monocytes and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: metabolism disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, melanocytes, kidney, blood cells, ovary and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATPase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of metabolism disorders, especially in fetal and tumor tissue growth.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene shares sequence homology with the immunoglobulin superfamily of proteins which are known to be important in immune response and immunity.

This gene is expressed primarily in stromal cells, colon cancer, lung, amygdala, melanocyte and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells, colon, lung, amygdala, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune system disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 86

The translation product of this gene shares sequence homology with transcription initiation factor eIF-4 gamma which is thought to be important in gene transcription.

This gene is expressed primarily in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to transcription initiation factor eIF-4 gamma indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene regulation in tumorigenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

The translation product of this gene shares sequence homology at low level in prolines with secreted basic proline-rich peptide II-2 which is thought to be important in protein structure or inhibiting hydroxyapatite formation in vitro.

This gene is expressed primarily in endometrial tumor and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: endometrial tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular/skeletal and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secreted basic proline-rich peptide II-2 indicate that polynucleotides and polypeptides corresponding to this gene are useful for inhibiting hydroxyapatite formation or establishing cell/tissue structure.

FEATURES OF PROTEIN ENCODED BY GENE NO: 88

This gene is expressed primarily in: amniotic cells induced with TNF in culture; and to a lesser extent in colon tissue from a patient with Crohn's Disease; parathyroid tumor; activated T-cells; cells of the human Caco-2 cell line; adenocarcinoma; colon; corpus colosum; fetal kidney; pancreas tumor; fetal brain; early stage brain, and anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system; e.g., tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., amniotic cells, colon, kidney, pancreas, parathyroid, brain and other tissue of the nervous system, blood cells, hematopoietic cells, liver, spleen, bone, testes and other reproductive tissue, brain and other tissue of the nervous system, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for modulating tumorigenesis and other immune system conditions such as disorders in immune response.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

This gene is expressed primarily in fetal liver/spleen and hematopoietic cells and to a lesser extent in brain, osteosarcoma, and testis tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, liver, spleen, bone, testes, and other reproductive tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

The translation product of this gene shares weak sequence homology with mouse Gcapi protein which is developmentally regulated in brain.

This gene is expressed primarily in infant and adult brain and fetal liver/spleen and to a lesser extent in smooth muscle, T cells, and a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, spleen, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and its homology to Gcapi protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosis of disorders in neuronal, hematopoietic, immune, and endocrine systems.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in brain and hematopoietic cells and to a lesser extent in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorder in nervous, hematopoietic, immune systems and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of disorders in the nervous, hematopoietic, and immune systems.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The translation product of this gene shares sequence homology with neuroendocrine-specific protein A which is thought to be important in neurologic systems.

This gene is expressed primarily in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neural disorders and degeneration disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central or peripheral nervous systems, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to neuroendocrine-specific protein A indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neural disorders and degeneration disease.

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 93

The translation product of this gene shares sequence homology with collagen-like protein and prolin-rich protein which are thought to be important in connective tissue function and tissue structure.

15

This gene is expressed primarily in fetal liver/spleen and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution and homology to collagen-like protein and proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for supporting brain and hematopoietic tissue function and diagnosis and treatment of disorders in these functions.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 94

This gene is expressed primarily in embryonic tissues and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

35

biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system (e.g., tumors), expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in brain tumor, placenta, and melanoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor or melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain or melanocytes, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, placenta, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that the translation product of this gene is useful in the diagnosis and treatment of brain tumors and melanoma.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

30

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

35

This gene is expressed primarily in fetal liver, and to a lesser extent in placenta and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of fetal liver or defects of glucose-regulated ATPase activities in tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune/hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, placenta, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to yeast SUR4 membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of defects of fetal liver or defects of glucose-regulated ATPase activities.

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

This gene is expressed primarily in fetal liver, brain, and amniotic fluid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the fetal immune system and adult brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune system and adult brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the fetal immune and hematopoietic systems since fetal liver is

the predominant organ responsible for hematopoiesis in the fetus. In addition, the gene product of this gene is thought to be useful for detecting certain neurological defects of the brain.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

The translation product of this gene shares sequence homology with an yolk protein precursor, Vitellogenin which is thought to be important in binding lipids such as phosvitin.

This gene is expressed primarily in amniotic cells and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in amniotic cells, fetal liver development and the fetal immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to vitellogenin indicate that the protein product of this clone is useful for treatment and diagnosis of defects in amniotic cells, fetal liver development and the fetal immune system.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in placenta, endometrial tumor, osteosarcoma and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumor of the endometrium or bone, and osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the obstetric system (e.g. placenta,

30

35

endometrium) and the bones, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endometrium, bone, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors and abnormalities of the endometrium, and the bones because of its abundance in the aforementioned tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene is expressed primarily in hepatocellular tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatocellular tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of hepatocellular cancer because of its abundant expression in this tissue.

FEATURES OF PROTEIN ENCODED BY GENE NO: 101

This gene is expressed primarily in Corpus Colosum, fetal lung and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the Corpus Colosum or defects of the fetal lung. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Corpus Colosum and brain in general, and fetal lung, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of the Corpus Colosum and brain in general, and defects of fetal lung.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 102

This gene is expressed primarily in T cells and stromal cells, and to a lesser extent in adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of T cell immunity and stromal cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, stromal cells, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of T cell immunity and stromal cell development because of its abundant expression in these tissues.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 103

This gene is expressed primarily in infant brain and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, especially brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the brain, especially in young children.

FEATURES OF PROTEIN ENCODED BY GENE NO: 105

This gene is expressed primarily in human osteoclastoma and to a lesser extent in human pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly osteoclastoma and pancreatic tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in transformed tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone and pancreas, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of some types of tumors, particularly pancreatic cancer and osteoclastoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 106

This gene is expressed primarily in fetal liver/spleen, and to a lesser extent in activated T-Cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis or treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 107

This gene is expressed primarily in human embryo and to a lesser extent in spleen and chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the diagnosis and treatment of leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene is expressed primarily in placenta, and to a lesser extent in early stage human brain and in lung.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: fetal developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in fetal and amniotic tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this is useful for production of growth factor(s) associated with fetal development. Preferred polypeptides comprise the full-length polypeptide shown in the sequence listing, truncated however, at the amino terminus and beginning with QTIE.

FEATURES OF PROTEIN ENCODED BY GENE NO: 109

This gene is expressed primarily in fetal spleen, and to a lesser extent in B-Cell lymphoma and T-Cell lymphoma.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the treatment and diagnosis of human lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 110

5 The translation product of this gene shares sequence homology with sarcoma amplified sequence (SAS), a tetraspan receptor which is thought to be important in malignant fibrous histiocytoma and liposarcoma.

This gene is expressed primarily in human osteoclastoma, and to a lesser extent in pineal gland and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: malignant fibrous histiocytoma and liposarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, pineal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to sarcoma amplified sequence (SAS) indicate that the protein product of this clone is useful for treatment of, osteosarcoma, malignant fibrous histiocytoma and liposarcoma and related cancers, particularly sarcomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 111

The translation product of this gene shares sequence homology with 6.8K proteolipid protein, mitochondrial - bovine.

30 This gene is expressed primarily in Wilm's tumor and to a lesser extent in cerebellum and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Wilm's tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the immune or renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to 6.8K proteolipid protein indicate that the protein product of this clone is useful for diagnostic and therapeutics associated with tumors, particularly Wilm's tumor disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 112

This gene is expressed primarily in embryonic tissue and to a lesser extent in osteoblasts, endothelial cells, macrophages (GM-CSF treated), and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, bone, endothelial cells, blood cells and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune disorders. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: MITDQLAIFANMLGVSLFLVLYHYVAVNNPKQE (SEQ ID NO: 636).

FEATURES OF PROTEIN ENCODED BY GENE NO: 113

This gene is expressed primarily in hepatocellular tumor, and to a lesser extent in fetal liver/spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors, particularly hepatocellular tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of tumors, particularly hepatocellular tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene exhibits a very high degree of sequence identity with the human Pig8 gene which is thought to be important in p53 mediated apoptosis. The sequence of this gene has since been published by Polyak and colleagues (Nature 389, 300-306 (1997)). In addition, the predicted translation product of this contig exhibits very high sequence homology with a murine gene denoted as EI24 which is also thought to be important in p53 mediated apoptosis.

This gene is expressed primarily in infant brain and activated T-cells and to a lesser extent in bone marrow, fetal liver, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and tissue damage by radiation and anti-cancer drugs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, bone marrow, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human Pig8 and murine E124 genes

indicate that polynucleotides and polypeptides corresponding to this gene are useful for preventing apoptosis in patients being treated with anti-oncogenic drugs such as etoposide, hydroperoxycyclophosphamide, and X-irradiation, since this protein product is upregulated in cells undergoing such treatment where p53 was overexpressed. It may also be useful in the treatment of hematopoietic disorders and in boosting numbers of hematopoietic stem cells by interfering with the apoptosis of progenitor cells. The mature polypeptide is predicted to comprise the following amino acid sequence:

10 EEMADSVKTFQLDARGIKDSTWIGICTSKLDARIQKREQRRRRRASSVLQRRRAQSIERKQES
EPRIVRFQCCAWNGGVWFSLLLFYRVFIPVLSQSVTARIUGDPSLHGDVWSWLEFFLTSIFSA
LWVLPFLVLSKVNAIWFQDIADLAFAEYSGRKPFPFSVKIADMLFNLLQALFLIQGMFVSL
FPHLVGQLVSLHMSLLYSLYCFEYRVFNKGIEMHQRLSNIERNWPYYFGCLPLAFLTAMQ
15 SSYISGCLFSILPFLFITSANEAKTPGKAYLFQLRLFSLVFLSNRLFKHTVYQLQSALSSSTSAEK
FSPHPSPAKLKATAGH (SEQ ID NO: 637). Accordingly, polypeptides comprising the foregoing amino acid sequence are provided as are polynucleotides encoded such polypeptides.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 115

This gene is expressed primarily in stromal cells and to a lesser extent in multiple sclerosis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: affecting the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of multiple sclerosis and other autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 116

This gene is expressed primarily in the gall bladder

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: gall stones or infection of the digestive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for possible prevention of digestive disorders where there may be a lack of digestive enzymes produced or in the detection and possible prevention of gall stones.

FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with dystrophin gene which is thought to be important in building and maintenance of muscles.

25 This gene is expressed primarily in placenta and to a lesser extent in fetal brain and fetal liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: muscular dystrophy. Duchenne and Becker's muscular dystrophies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal muscle system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, muscle, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the dystrophin gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diseases related the degenerative myopathies that are characterized by the weakness and atrophy of muscles without neural degradation, such as Duchenne and Becker's muscular dystrophies.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 118

This gene is expressed primarily in olfactory tissue and to a lesser extent in cartilage.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: connective tissue diseases; chondrosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the connective tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue and cartilage, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for tumors of connective tissues, osteoarthritis and the treatment and diagnosis of chondrosarcoma.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 119

This gene is expressed primarily in Activated Neutrophils and to a lesser extent in fetal spleen, and CD34 positive cells from cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergies, defects in hematopoiesis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoiesis system the, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for reducing the allergic effects felt by allergy suffers by neutralizing the activity of the immune system, especially since neutrophils are abundant in persons suffering from allergies and other inflammatory conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 120

The translation product of this gene shares sequence homology with poly A binding protein II which is thought to be important in RNA binding for transcription of RNA to DNA

This gene is expressed primarily in colon and to a lesser extent in brain and immune system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: colon cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, tissue and cells of the immune system, and brain or other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to poly A binding protein II indicate that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of colon cancer and other disorders of the digestive system..

FEATURES OF PROTEIN ENCODED BY GENE NO: 121

The translation product of this gene shares sequence homology with thymidine diphosphoglucose 4.6 dehydratase which is thought to be important in the metabolism of sugar.

5 This gene is expressed primarily in fetal liver and spleen and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diabetes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution and homology to thymidine diphosphoglucose 4.6 dehydratase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of persons with diabetes since it appears that this protein is needed in the metabolism of sugar in to its more basic components.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with ceruloplasmin which is thought to be important in the metabolism and transport of iron and copper. Ceruloplasmin also contains domains with homology to clotting factors V and VIII. Defects in the circulating levels of ceruloplasmin (aceruloplasminemia) have been associated with certain disease conditions such as Wilson disease, and the accompanying hepatolenticular degeneration.

30 This gene is expressed primarily in brain and retina and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases marked by defects in iron metabolism; aceruloplasminemia not characterized by defects in the

5 known ceruloplasmin gene locus; nonclassical Wilson disease; movement disorders; and tumors derived from a brain tissue origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, retina, and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, retinal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to ceruloplasmin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of patients with aceruloplasminemia, or other defects in iron and/or copper metabolism. Mutations in this locus could also be diagnostic for patients currently experiencing or predicted to experience aceruloplasminemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 123

20 This gene is expressed primarily in brain and B cell lymphoma and to a lesser extent in fetal liver and spleen.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: B cell lymphoma; tumors and diseases of the brain and/or spleen; hematopoietic defects. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in neuronal,

hematopoietic, and immune systems. It could potentially be useful for neurodegenerative disorders and neuronal and/or hematopoietic cell survival or proliferation.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in osteoclastoma, dermatofibrosarcoma, and B cell lymphoma and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer in particular osteoclastoma, dermatofibrosarcoma, and B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, immune, and circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, epidermis, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers and lymphoma, osteoporosis; and the control of cell proliferation and/or differentiation.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in immune tissues and hematopoietic cells, particularly in activated T cells and neutrophils, spleen, and fetal liver, and to a lesser extent in infant adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in T cell activation; hematopoietic disorders; tumors of a hematopoietic and/or adrenal gland origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and/or endocrine systems, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues and cell types (e.g., cells and tissues of the immune system, hematopoietic cells, blood cells, liver, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and/or hematopoietic disorders; diseases related to proliferation and/or differentiation of hematopoietic cells; defects in T cell and neutrophil activation and responsiveness; and endocrine and/or metabolic disorders, particularly of early childhood.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in placenta and endothelial cells and to a lesser extent in melanocytes and embryonic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell origin; angiogenesis associated with tumor development and metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, melanocytes, and embryonic tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of developmental disorders; inhibition of angiogenesis; and vascular patterning.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in endothelial cells and hematopoietic tissues, including spleen, tonsils, leukocytes, and both B- and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell and/or hematopoietic origin; leukemias and lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, hematopoietic cells, spleen, tonsils, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the manipulation of angiogenesis; the differentiation and morphogenesis of endothelial cells; the proliferation and/or differentiation of hematopoietic cells; and the commitment of hematopoietic cells to distinct cell lineages.

FEATURES OF PROTEIN ENCODED BY GENE NO: 128

This gene is expressed primarily in kidney medulla and to a lesser extent in spleen from chronic myelogenous leukemia patients, prostate cancer, and some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a kidney origin; chronic myelogenous leukemia; prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, spleen, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of kidney disorders and cancer, particularly chronic myelogenous leukemia and prostate cancer. It may also be useful for the enhancement of kidney tubule regeneration in the treatment of acute renal failure.

FEATURES OF PROTEIN ENCODED BY GENE NO: 129

This gene is expressed primarily in adult and infant brain and to a lesser extent in mesenchymal or fibroblast cells, as well as tissues with a mesenchymal origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and of mesenchymal cells and tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; and fibrosis, based upon the expression of this gene within those tissues. Fibrosis is considered as mesenchymal cells and fibroblasts are the primary cellular targets involved in this pathological condition.

FEATURES OF PROTEIN ENCODED BY GENE NO: 130

This gene is expressed primarily in hepatocellular cancer and to a lesser extent in fetal tissues as well as testes tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, fetal tissue, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 131

This gene is expressed only in infant early brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: development and diseases of the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the brain in children and in treating nervous system disorders such as Alzheimer's disease, schizophrenia, dementia, depression, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in brain and to a lesser extent in glioblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Alzheimer's disease,

schizophrenia, depression, mania, and dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating brain disorders such as Alzheimer's disease, schizophrenia, depression, mania, and dementia.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 133

The translation product of this gene shares sequence homology with ribitol dehydrogenase of bacteria which is thought to be important in metabolism of sugars.

This gene is expressed primarily in macrophage and to a lesser extent in T-cell lymphoma and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tissue destruction in inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribitol dehydrogenase indicate that polynucleotides and polypeptides corresponding to this gene are useful for altering macrophage metabolism in diseases such as inflammation where macrophages are causing excess tissue destruction.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 134

This gene is expressed primarily in pancreatic tumor and to a lesser extent in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and connective tissue systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pancreas, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing various cancers.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 135

This gene is expressed primarily in T cell lines such as Raji and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune system disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing inflammatory diseases

such as rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, as well as neutropenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 136

The translation product of this gene shares high sequence homology with SAR1 subfamily of GTP-binding proteins which is thought to be important in vesicular transport in mammalian cells.

This gene is expressed primarily in serum-stimulated smooth muscle cells and to a lesser extent in a T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases affecting vesicular transport. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP-binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene therapy in treating the large number of diseases involved in defective vesicular transport within cells..

FEATURES OF PROTEIN ENCODED BY GENE NO: 137

The translation product of this gene shares sequence homology with a protein found in *C. elegans* cosmid F25B5.

This gene is expressed primarily in a fetal tissues and to a lesser extent in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development, especially of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal pulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g.; fetal tissue, pulmonary tissue, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

10 corresponding to this gene are useful for treatment and diagnosis of diseases affecting the pulmonary system, such as emphysema.

FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene is expressed primarily in gall bladder and to a lesser extent in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: digestive system disease and gall bladder problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

30 corresponding to this gene are useful for treating diseases of the digestive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene is expressed primarily in placenta and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developing tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing abnormal fetal development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene is expressed primarily in smooth muscle and to a lesser extent in ovary, prostate cancer, and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hypertension and atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, ovary and other reproductive tissue, prostate, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the circulatory system, such as hypertension, atherosclerosis, etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 141

This gene is expressed primarily in fetal spleen and to a lesser extent in placenta and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and other diseases affecting blood cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen, placenta, bone marrow, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the generation of red and white blood cells and for the diagnosis of disease of these cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 142

The predicted translation product of this contig is a human homolog of the murine tetracycline/sugar transporter molecule recently reported by Matsuo and colleagues (Biochem. Biophys. Res. Commun. 238 (1), 126-129 (1997)).

This gene is expressed primarily in synovium and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: rheumatoid arthritis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and lymphatic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, such as rheumatoid arthritis, leukemia, neutropenia, inflammatory bowel disease, psoriasis, sepsis, and the like.

FEATURES OF PROTEIN ENCODED BY GENE NO: 143

This gene is expressed primarily in placenta and to a lesser extent in melanocyte, fetal liver and spleen, and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal early development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, melanocytes, liver, spleen, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of abnormal early development phenomena and diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and blood systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in hematopoiesis and bone marrow regeneration as it is most abundant in fetal tissues responsible for the generation of hematopoietic cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 145

The translation product of this gene shares sequence homology with protein tyrosine phosphatase which is thought to be important in transducing signal to activate cells such as T cell, B cell and other cell types.

This gene is expressed primarily in T cells and tissues in early stages of development and to a lesser extent in cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic and fetal tissue, undifferentiated cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the protein tyrosine phosphatase family indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating the immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 146

This gene is expressed primarily in T cell and to a lesser extent in B cell, macrophages and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating the immune system therefore can be used in treating diseases such as autoimmune diseases and cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 147

This gene is expressed primarily in placenta and to a lesser extent in endothelial cells, testis tumor, ovarian cancer, uterine cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, testis and ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This sequence has significant homology to mouse torsin A. Recently, another group cloned the human Torsin A gene. (See, Accession No. 2358279; see also Nature Genet. 17, 40-48 (1997).)

This gene is expressed primarily in osteoclastoma, T-cell, and placenta and to a lesser extent in fetal lung, fetal liver, fetal brain, adult brain and tumor tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disease conditions in hematopoiesis and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, bone, placenta, lung, liver, and brain and other tissues of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating blood related diseases such as deficiencies in red blood cell, white blood cell, platelet and other hematopoiesis cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 149

This gene is expressed primarily in T cell, prostate and prostate cancer, endothelial cells and to a lesser extent in monocyte, dendritic cell, bone marrow, salivary gland, colon cancer, stomach cancer, pancreatic tumor, uterine cancer, fetal spleen and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, prostate, endothelial cells, dendritic cells, bone marrow, salivary gland, colon, stomach, pancreas, uterus, spleen and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 150

This gene was recently cloned by another group, calling it cIF3-p66. (See Accession No. 2351378.) This gene plays a role in RNA binding and macromolecular assembly, and therefore, any mutations in this gene would likely result in a diseased phenotype. Preferred polypeptide fragments comprise the amino acid sequence:

MAKFMTPVIQDNPDSGWCPCAVPEQFRDMPYQPFSGKDRLGKLVADWTOATTQDKRYTNKYSS
QFGGGSQAYFHEEDESFLVDARTQKTAYQRNRMFAQRNLRRDKDRNNMLQPNLQILP
KSAKQKERERLQKKFQKQVGRQKWDQSKQPRDSSVEVRSDWVYKEEMDPQLMKM
LEVSEPDIECCGALYYDYDKAFDRITTRSEKPLRXKKRIFHTVTTDDPVIRKLAKTQNNVFA
AILATLMCTRSYVSWDITVQRYGSKLFFDKRDNDFDILLTVSETANEPQDEGNFSRPNRL
AMEATYNHNSQQCLRMGKERYNFPNPFVEDDDMDKNELASVAYRYSGKLGGDDIDLVR
EHDGVMTOANGEVSEFINIKTLNWDNRHNGVDWRQKLDQSRGAVIA TELKNNSYKLARWTC
CALLAGSEYLKLOYSVRYHVKDSSRHVLGTQQFKPNEFASQINLSVENAWGLRCVIDICMKL
ESGKYLKDPNKQVIRVYSLPDGTSS (SEQ ID NO: 638), as well as N-terminal and C-terminal deletions of this polypeptide fragment.

This gene is expressed primarily in T cell, bone marrow, embryo and endothelial cells and to a lesser extent in testis tumor and endometrial tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders and cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 151

This gene is expressed primarily in testis and to a lesser extent in T cell, spinal cord, placenta, neutrophil and monocyte.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive tissue, blood cells, tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating immune and reproductive functions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 152

The translation product of this gene shares sequence homology with tyrosyl-tRNA synthetase which is thought to be important in cell growth.

This gene is expressed primarily in brain, liver, keratinocytes, tonsils, and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, keratinocytes, tonsils, heart expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissues of the nervous system, liver, keratinocytes, tonsils and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to tyrosyl-tRNA synthetase indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

FEATURES OF PROTEIN ENCODED BY GENE NO: 153

This gene is homologous to the *Drosophila* transcriptional regulator *dre4*. (See Accession No. 2511745.) *Dre4* is a gene required for steroidogenesis in *Drosophila melanogaster* and encodes a developmentally expressed homologue of the yeast transcriptional regulator CDC68. Preferred polypeptide fragments comprise the amino acid sequence: KKRRTDVQFYTEVGEITTDLGKQHQMHDRDDLVAEQMERMRHKLKTAFLKFIKVEALTKEELEFVPRDLGFNGAPYRSTCLLQPTSSALVNA TEWPPFVVTLDEVELIHFXR VQFHLKNFDMVIVYKDYSKKYTMINAIPVASLDPIKEWLNSCDLKYTEGVQSLNWTGKMTYVD DPEQFTEQGQWSFL (SEQ ID NO: 639), as well as N-terminal and C-terminal deletions of this fragments. Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in fetal liver, spleen, placenta, lung, T cell, thyroid, testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor, heart and liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal liver, spleen, placenta, lung, T cell, thyroid, testes expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, lung, blood cells, thyroid, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed primarily in brain and to a lesser extent in fetal heart, testis, spleen, lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: heart, liver and spleen diseases, immunological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, fetal heart, testis, spleen, lung expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, heart, testes and other reproductive tissue, spleen, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 155

Activation of T cells through the T cell antigen receptor (TCR) results in the rapid tyrosine phosphorylation of a number of cellular proteins, one of the earliest being a 100 kDa protein. This gene is the human equivalent of murine valosin containing protein (VCP). VCP is a member of a family of ATP binding, homo-oligomeric proteins, and the mammalian homolog of *Saccharomyces cerevisiae cdc48p*, a protein essential to the completion of mitosis in yeast. Both endogenous and expressed murine VCP are tyrosine phosphorylated in response to T cell activation. Thus we have identified a novel component of the TCR mediated tyrosine kinase activation pathway that may provide a link between TCR activation and cell cycle control.

25

This gene is expressed primarily in brain, liver, spleen, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, spleen, placenta expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, spleen, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

35

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to VCR indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating cancer.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with rat growth response protein which is thought to be important in cell growth. A group recently cloned the human homolog of this gene, calling it insulin induced protein 1. (See Accession No. 2358269, see also, Genomics 43 (3), 278-284 (1997).) Preferred polypeptide fragments comprise the amino acid sequence: RSGLGIGITIAFLATLTQPR LVYNGVYQYTSDFLYIRSWLPCIFFSGVTVGNIGROLAMGVPEKPHSD (SEQ ID NO: 640), as well as N-terminal and C-terminal deletions of this polypeptide fragment. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

15

This gene is expressed primarily in brain, liver, placenta, heart, spleen, lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, placenta, heart, spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, placenta, heart, spleen, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution and homology to growth-response protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

30

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 157

This gene is expressed primarily in Glioblastoma, endometrial tumor, lymphoma and pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Glioblastoma, Endometrial tumor, lymphoma and pancreas tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, lymphoid tissue, pancreas, and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 158

The translation product of this gene shares sequence homology with IGE receptor which is thought to be important in allergy and asthma.

This gene is expressed primarily in T cell, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergy and asthma and other immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to IgE receptor indicate that

polynucleotides and polypeptides corresponding to this gene are useful for allergy and asthma.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 159

The translation product of this gene shares sequence homology with immunoglobulin heavy chain which is thought to be important in immune response to the antigen.

This gene is expressed primarily in activated neutrophil and to a lesser extent in activated T cell, monocyte and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: infection, inflammation and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin heavy chain variable region indicate that polynucleotides and polypeptides corresponding to this gene are useful for making the ligand to block specific antigen which cause certain disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with mouse X inactive specific transcript protein which is thought to be important in X chromosome inactivation.

This gene is expressed primarily in HSA172 cell and to a lesser extent in normal ovary tissue, ovarian cancer, frontal cortex and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian tumor, schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

5 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to X inactive specific transcript protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive system tumors and CNS tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 161

15 This gene is expressed primarily in adipose cell and to a lesser extent in liver and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity and liver disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose cell, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose cells, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of obesity and liver disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 162

35 The translation product of this gene shares sequence homology with yeast ubiquitin activating enzyme homolog which is thought to be important in protein posttranslation processing.

This gene is expressed primarily in stromal cell and to a lesser extent in retina, H. Atrophic Endometrium, colon carcinoma and myeloid progenitor cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development, neuronal growth disorders and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal cells, endometrium, colon, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ubiquitin-activating enzyme homolog indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of some type of tumors, fucosidosis and neuronal growth disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in primary breast cancer and hemangiopericytoma and to a lesser extent in adult brain and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer, leukemia and cerebellum disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis or treatment of various tumors and disease involved in neural system.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 164

The translation product of this gene shares sequence homology with proline rich proteins. Recently, another group has also cloned this gene, calling it CD84 leukocyte antigen, a new member of the Ig superfamily. (See Accession No. U82988, see also, Blood 90 (6), 2398-2405 (1997).)

10 This gene is expressed primarily in Weizmann olfactory tissue and osteoclastoma and to a lesser extent in anergic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteitis and immune disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue, bone, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to the Ig superfamily indicate that the protein product of this clone is useful for treatment of osteoporosis, autoimmune disease, and other immune disorders.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 165

This gene is expressed primarily in atrophic endometrium and colon cancer and to a lesser extent in some fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, colon, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors, specifically endometrium and colon tumors.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human primary breast cancer and to a lesser extent in activated monocyte. Although the predicted signal sequence is identified in Table 1, other upstream sequences are also relevant. Preferred polypeptide fragments comprise the amino acid sequence: VTQPKHLASMGSGVEIFSFYYPWELAXXPXVRLSWRRGHFHG QSFYSTRPPSHKDYVNRLFLNWTGQESGFLRISNLRKEDQSVYPCRVELDTRSG (SEQ ID NO: 641), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of breast cancer.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 167

This gene is expressed primarily in fetal tissues and to a lesser extent in adult lung. This gene has also been mapped to chromosomal location 9q34, and thus, can be used as a marker for linkage analysis for chromosome 9.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryo tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 168

The translation product of this gene shares sequence homology with Ig Heavy Chain which is thought to be important in immune response.

This gene is expressed primarily in prostate cancer tissue specifically

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate cancer.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The translation product of this gene shares sequence homology with cytosolic acyl coenzyme-A hydrolase, which is thought to be important in neuron-specific fatty acid metabolism. The gene represented by this contig has since been published by Hajra and colleagues (GenBank Accession No. U91316).

This gene is expressed primarily in human pituitary gland and to a lesser extent in colorectal cancer tissue. This gene has also been observed in the LNCAP cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hyperlipidemias of familial and/or idiopathic origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly blood, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to rat cytosolic acyl coenzyme-A

hydrolase indicate that polynucleotides and polypeptides corresponding to this gene are useful for the detection or treatment of hyperlipidemia disease states by virtue of the ability of specific drugs to activate the enzyme.

FEATURES OF PROTEIN ENCODED BY GENE NO: 170

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene which is thought to be important in organism development.

This gene is expressed primarily in human synovial sarcoma tissue, bone marrow, and to a lesser extent in human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of bone, specifically synovial sarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, connective tissues and possibly immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, bone marrow, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to *Caenorhabditis elegans* indicate that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic and/or therapeutic modality directed at the detection and/or treatment of connective tissue sarcomas or other related bone diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 171

The translation product of this gene shares sequence homology with beta1-6GlcNAc transferase which is thought to be important in the transfer and metabolism of beta1-6, N-acetylglucosamine. This gene product has previously been shown to suppress melanoma lung metastasis in both syngeneic and nude mice, decreased invasiveness into the matrigel, and inhibition of cell attachment to collagen and laminin without affecting cell growth.

This gene is expressed primarily in human testes and prostate tissues, and to a lesser extent in kidney, medulla, and pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, prostate, kidney, pancreas, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to beta1-6GlcNAc transferase indicate that the protein product of this clone is useful for the development of diagnostic and/or therapeutic modalities directed at the detection and/or treatment of cancer, the metastasis of malignant tissue or cells. Defects in this potentially secreted enzyme may play a role in metastasis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in fetal spleen and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders, Wilm's tumor disease, hepatic disorders, and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and identification of fetal defects along with correcting diseases that affect hematopoiesis and the immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 173

The translation product of this gene shares sequence homology with ret II oncogene which is thought to be important in Hirschsprung disease and many types of cancers.

This gene is expressed in multiple tissues including the lymphatic system, brain, and thyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Hirschsprung disease and multiple cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly higher or lower levels central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, thyroid, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ret II oncogene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of various cancers. It would also be useful for the diagnosis and treatment of Hirschsprung disease. Preferred polypeptides of the invention comprise the amino acid sequence: MEAQQVNEAESAREQLQXLHDQIAGQKASKQEQLETLERLKQEFHYTEEDLY RTKNTLSRIKDRDEEQIKLRNQLTNKTLSNSQSELENRLHQLJETLTQKQTMLESISTEKNLS VFQLERLEQQMNSAGSSSSNGSSINMSGIDNGEGTGLRNVPVLFNDTETNLNAGMYGKYRKAAS SIDQFSIRLGIPLRRYPPIARVFVHYIMALLHLWYVMIVLLTYPTEM HHDPQYCK (SEQ ID NO: 642).

FEATURES OF PROTEIN ENCODED BY GENE NO: 174

The translation product of this gene shares sequence homology with testis enhanced gene transcript which is thought to be important in regulation of human development.

This gene is expressed primarily in infant brain and to a lesser extent in a variety of other tissues and cell types, including the prostate, testes, monocytes, macrophages, dendritic cells, keratinocytes, and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological, developmental, immune and inflammation disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, prostate, testes and other reproductive tissue, blood cells, keratinocytes, and adipocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to testis enhanced gene transcript indicate that the protein product of this clone is useful for diagnosis and treatment of disorders involving the developing brain and the immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 175

This gene is expressed primarily in prostate and to a lesser extent in various other tissues, including placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, especially of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of prostate disorders and cancer. It may also be useful for the diagnosis and treatment of endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with *Saccharomyces cerevisiae* YNT20 gene which is thought to be important in mitochondrial function.

This gene is expressed at a particularly high level in muscle tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases related to such tissues and cell types including: muscle wasting diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., muscle and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the YNT20 gene indicate that this protein is useful for treatment and detection of neuromuscular diseases caused by loss of mitochondrial function. For example this gene or its protein product could be used in replacement therapy for such diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 177

This gene is expressed primarily in the brain and to a lesser extent in kidney, placenta, smooth muscle, heart and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuromuscular diseases, degenerative diseases of the central nervous system, and heart disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, central nervous system, and heart, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, kidney, placenta, muscle, heart and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

This gene or its protein product could also be used for replacement therapy for the above mentioned diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 178

The translation product of this gene shares sequence homology with caldesmon which is thought to be important in the cellular response to changes in glucose levels.

This gene is expressed primarily in multiple tissues including brain and retina. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: central nervous system disorders and retinopathy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the CNS disorders and retinopathy, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and retinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to caldesmon indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of retinopathies.

FEATURES OF PROTEIN ENCODED BY GENE NO: 179

The translation product of this gene shares sequence homology with mouse fibrosin protein which is thought to be important in regulation of fibrinogenesis in certain chronic inflammatory diseases.

This gene is expressed primarily in amniotic cells and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of breast cancer and abnormal embryo development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to fibrosin indicate that the protein product of this clone is useful for treatment of breast cancer. This gene or its protein product could be used in replacement therapy for breast cancer. In addition the protein product of this gene is useful in the treatment of chronic inflammatory diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 180

This gene is expressed several infant tissues including brain and liver and various adult tissues including brain, lung, liver, testes, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, brain cancer, lung cancer, liver cancer and cancers of the reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, liver, testes and other reproductive tissue, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene product indicates that the protein product of this clone is involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in activated monocytes and to a lesser extent in melanocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of immune system diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, melanocytes, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 182

This gene is expressed primarily in placenta and several tumors of various tissue origin and to a lesser extent in normal tissues including liver, lung, brain, and skin,

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers of all kinds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, respiratory system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, lung, brain and other tissues of the nervous system, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The high expression of this gene in multiple tumors indicates that the protein product of the clone may be involved in cell growth control and therefore would be useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

FEATURES OF PROTEIN ENCODED BY GENE NO: 183

The translation product of this gene shares sequence homology with the mouse Ndr1 gene which is thought to be important in cancer progression.

This gene is expressed multiple cell types and tissues including brain, lung, kidney, bone marrow, liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, kidney, bone marrow, liver and spleen, and cancerous and wounded

(tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to Ndr1 gene, which is thought to be involved in cancer progression, indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 184

This gene is expressed primarily in early stage human brain and liver and to a lesser extent in several other fetal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain and liver cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 185

This gene is expressed primarily in infant and embryonic brain.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of degenerative nervous system disorders and brain cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 186

15 This gene is expressed primarily in multiple tissues including placenta, fetal lung, fetal liver, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers including liver, brain and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, pulmonary system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, lung, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	11	582	1	582	177	177	313	1	18	19	22
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	197	1020	296	830	442	442	499	1	18	19	22
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	12	465	1	465	81	81	314	1	30	31	128
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	198	524	229	343		196	500	1	20	21	33
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	13	474	1	474	1	1	315	1	24	25	28
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	199	332	1	319	35	35	501	1	24	25	28

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
4	HCUFQ22	97897 02/26/97 209043 05/15/97	ZAP Express	14	314	1	298	122	122	316	1	34	35	64
5	HCUFV01	97897 02/26/97 209043 05/15/97	ZAP Express	15	613	1	613	30	30	317	1	18	19	21
6	HCUGA50	97897 02/26/97 209043 05/15/97	ZAP Express	16	356	1	356	239	239	318	1	22	23	39
7	HCUIM14	97897 02/26/97 209043 05/15/97	ZAP Express	17	414	185	414	278	278	319	1	26	27	33
8	HLDOU93	97897 02/26/97 209043 05/15/97	pCMVSPORT 3.0	18	469	1	469	77	77	320	1	44	45	88
9	HEIAX07	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	19	550	1	550	129	129	321	1	21	22	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
9	HELAX07	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	200	376	9	376		1	502	1	8	9	15
10	HSAXR76	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	20	741	55	741	190	190	322	1			27
11	HNGJJ68	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	21	991	1	991	62	62	323	1	30	31	64
11	HNGJJ68	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	201	1192	253	1137		409	503	1			19
12	HCFAW04	97897 02/26/97 209043 05/15/97	pSport1	22	653	1	653	64	64	324	1	30	31	196
12	HCFAW04	97897 02/26/97 209043 05/15/97	pSport1	202	589	1	513	109	109	504	1			29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	23	1486	596	1418	102	102	325	1	54	55	252
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	203	847	1	839	87	87	505	1	30	31	75
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	204	852	75	850		690	506	1			10
13	HTXEF04	209235 09/04/97	Uni-ZAP XR	205	1354	54	1354	100	100	507	1	33	34	207
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	24	2323	1017	2059	1242	1242	326	1	21	22	68
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	206	1378	113	1226	303	303	508	1	25	26	36
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	25	683	1	683	304	304	327	1	30	31	84

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	207	1166	281	884	567	567	509	1	18	19	19
16	HHFFL33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	26	2036	14	1959	214	214	328	1	20	21	36
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	27	717	1	717	70	70	329	1	30	31	63
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	208	697	2	697	33	33	510	1	31	32	32
18	HMDAE90	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	28	495	1	495	39	39	330	1	24	25	35
19	HOUAW01	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	29	556	1	556	116	116	331	1	19	20	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
20	HBIAE44	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	30	434	1	434	78	78	332	1	35	36	40
21	HCFME41	97897 02/26/97 209043 05/15/97	pSport1	31	715	1	715	87	87	333	1	30	31	111
21	HCFME41	97897 02/26/97 209043 05/15/97	pSport1	209	932	274	932	387	387	511	1	27	28	28
22	HOGCO71	97897 02/26/97 209043 05/15/97	pCMVSPORT 2.0	32	486	1	486	137	137	334	1	21	22	106
23	HOSEX08	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	33	725	1	725	436	436	335	1	30	31	50
23	HOSEX08	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	210	661	1	647	81	81	512	1	25	26	26

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
24	HSKNJ72	97897 02/26/97 209043 05/15/97	pBluescript	34	437	1	437	85	85	336	1	30	31	48
25	HEBEB69	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	35	943	1	943	196	196	337	1	30	31	41
25	HEBEB69	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	211	592	1	534	72	72	513	1	24	25	33
26	HE6EH18	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	36	604	1	604	375	375	338	1	20	21	76
26	HE6EH18	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	212	938	1	509		17	514	1	30	31	47
27	HSAUZ47	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	37	349	1	349		201	339	1	20	21	31

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
28	HSSDM73	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	38	672	1	672	22	22	340	1	38	39	42
29	HBMVK68	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	39	1908	135	1908	309	309	341	1	20	21	26
30	HMKDC66	97898 02/26/97 209044 05/15/97	pSport1	40	458	93	458	147	147	342	1	24	25	26
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	41	1153	500	1153	427	427	343	1	30	31	157
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	213	1079	502	896		739	515	1	23	24	43
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	42	1983	1092	1983	27	27	344	1	11	12	520

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	214	3791	2757	3357		2030	516	1			3
33	HTOJN06	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	43	1406	1	695		19	345	1	19	20	39
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	44	1391	851	1153	74	74	346	1	30	31	234
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	215	1334	822	1036		638	517	1	18	19	174
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	45	1569	768	1569	14	14	347	1	19	20	169
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	216	1511	770	1404	844	844	518	1	32	33	43

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
36	HHPBD40	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	46	1924	1	1681	62	62	348	1	19	20	43
37	HOVCL83	97898 02/26/97 209044 05/15/97	pSport1	47	475	252	396	141	141	349	1	37	38	78
38	HBCAY62	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	48	346	1	346	61	61	350	1	19	20	24
39	HBICM48	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	49	1366	882	1300	177	177	351	1	30	31	274
39	HBICM48	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	217	642	192	581		448	519	1			13
40	HLTCL35	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	50	1405	110	1404	61	61	352	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
40	HLTCL35	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	218	1241	1	1241	172	172	520	1	21	22	30
41	HLHCK50	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	51	504	207	485	222	222	353	1			3
42	HRSAN45	97899 02/26/97 209045 05/15/97	ZAP Express	52	777	1	214	113	113	354	1	24	25	52
43	HSNBB14	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	53	602	1	419	41	41	355	1	59	60	132
43	HSNBB14	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	219	1080	186	686	399	399	521	1	26	27	47
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	54	1749	222	1749	166	166	356	1	30	31	204

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	220	1258	149	1190	254	254	522	1	18	19	26
45	HSKDK47	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	55	1896	596	1614	650	650	357	1	33	34	47
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	56	1753	555	1753	414	414	358	1	18	19	73
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	221	1693	554	1693		526	523	1	25	26	58
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	57	1220	690	1024	128	128	359	1	30	31	102
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	222	1196	712	1163		1097	524	1			19

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
48	HFCA174	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	58	1049	362	1049	335	335	360	1	33	34	48
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	59	1776	854	1737	189	189	361	1	30	31	179
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	223	1791	979	1791	1164	1164	525	1	18	19	40
50	HLFBC91	97899 02/26/97 209045 05/15/97	pBluescript SK-	60	443	1	443	164	164	362	1	21	22	25
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	61	2888	1909	2888	90	90	363	1	30	31	224
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	224	2517	1597	2517	1953	1953	526	1	18	19	57

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
52	HPRCE95	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	62	1851	1568	1736	139	139	364	1	30	31	349
52	HPRCE95	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	225	2424	299	2309		530	527	1	17	18	21
53	HHTLC66	97899 02/26/97 209045 05/15/97	ZAP Express	63	3542	883	3492	964	964	365	1	25	26	467
54	HMADJ02	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	64	883	237	883	229	229	366	1	30	31	152
54	HMADJ02	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	226	1080	242	1033	436	436	528	1	24	25	39
55	HPRCU93	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	65	1541	1	1541	236	236	367	1	30	31	373

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
55	HPRCU93	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	227	1336	4	1336	946	946	529	1	25	26	128
56	HSAXS65	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	66	732	41	698	163	163	368	1	18	19	83
56	HSAXS65	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	228	2043	1133	1756	1262	1262	530	1	20	21	82
57	HKTAG35	209011 04/28/97	Uni-ZAP XR	67	629	1	629	264	264	369	1			21
57	HMEFX42	97899 02/26/97 209045 05/15/97	Lambda ZAP II	229	540	25	536	227	227	531	1			20
58	HHFHN61	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	68	1751	375	1751	95	95	370	1	19	20	227
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	69	508	1	508	22	22	371	1	30	31	79

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	230	448	9	448		1	532	1	22	23	75
60	HHGCM20	97899 02/26/97 209045 05/15/97	Lambda ZAP II	70	245	1	245	93	93	372	1	1	2	51
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	71	361	1	361	1	1	373	1	30	31	61
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	231	407	1	407	210	210	533	1	17	18	60
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	72	713	8	713	169	169	374	1	30	31	40
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	232	830	190	580	329	329	534	1	28	29	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	73	862	1	862	67	67	375	1	30	31	44
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	233	932	138	905	287	287	535	1			2
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	74	4602	4162	4525	730	730	376	1	30	31	203
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	234	2786	2406	2739	2577	2577	536	1	22	23	36
65	HSGBA84	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	75	1255	1	1195	112	112	377	1	28	29	29
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	76	475	1	475	13	13	378	1	30	31	136

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	235	458	1	458	26	26	537	1			14
67	HTGCP16	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	77	465	25	299	74	74	379	1	33	34	41
68	HKIXR69	97900 02/26/97 209046 05/15/97	pBluescript	78	1907	1627	1730	26	26	380	1	30	31	468
68	HKIXR69	97900 02/26/97 209046 05/15/97	pBluescript	236	591	1	444	251	251	538	1			18
69	HETGJ09	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	79	1168	136	1168	267	267	381	1	20	21	29
70	HOBNC61	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	80	1285	132	1285	292	292	382	1	27	28	29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
71	HFFAH94	97900 02/26/97 209046 05/15/97	Lambda ZAP II	81	1290	768	1054	701	701	383	1	21	22	138
72	HBLAI95	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	82	684	1	684	119	119	384	1	30	31	74
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	83	2024	1609	1953	200	200	385	1	30	31	521
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	237	1286	391	959		1204	539	1	9	10	11
74	HEBEG68	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	84	931	14	537	85	85	386	1	25	26	137
75	HBLAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	85	825	59	802	66	66	387	1	30	31	186

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
75	HBLAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	238	734	1	734	1	1	540	1	37	38	108
75	HBLAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	239	809	80	794		294	541	1	15	16	106
76	HTXDU73	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	86	1238	36	918	17	17	388	1			1
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	87	1460	9	1458	166	166	389	1	53	54	299
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	240	2201	841	2080	507	507	542	1	43	44	136
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	241	1661	311	1520	390	390	543	1	35	36	424

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
78	HTEIY30	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	88	1395	567	1395	639	639	390	1	36	37	49
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	89	1186	352	1186	540	540	391	1	49	50	61
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	242	1146	329	1146	564	564	544	1	21	22	39
80	HPMFL27	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	90	1821	1203	1614	1503	1503	392	1	30	31	79
81	HMWDN32	97900 02/26/97 209046 05/15/97	Uni-Zap XR	91	862	253	862	359	359	393	1	32	33	36
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	92	696	349	696	98	98	394	1	30	31	180

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	243	1350	265	1230	348	348	545	1	32	33	58
83	HHFFW36	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	93	1886	1	1759	197	197	395	1			21
84	HE2PL77	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	94	1774	742	1772	785	785	396	1	21	22	60
85	HSDFV29	209076 05/22/97	Uni-ZAP XR	95	2503	1	1648	206	206	397	1	32	33	152
85	HCQAV53	97901 02/26/97 209047 05/15/97	Lambda ZAP II	244	1529	72	911	191	191	546	1	20	21	33
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	96	2801	418	2801	234	234	398	1	30	31	480
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	245	1537	1	1537	125	125	547	1	21	22	367

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
87	HLHDR57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	97	1631	916	1631	1	1	399	1	1	2	423
88	HAUAV32	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	98	504	26	504	197	197	400	1	23	24	78
88	HAUAV32	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	246	506	1	499	183	183	548	1	32	33	77
89	HNEBI60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	99	1416	145	1416	456	456	401	1	18	19	74
89	HNEBI60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	247	1348	84	1348	363	363	549	1	21	22	47
90	HSJC116	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	100	2847	1	2847		2	402	1			20

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
91	HTSEL31	97901 02/26/97 209047 05/15/97	pBluescript	101	1394	608	1346	602	602	403	1	23	24	87
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	102	794	1	794	518	518	404	1	30	31	92
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	248	1766	42	1766	356	356	550	1	30	31	168
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	249	2664	47	1708		147	551	1	18	19	124
93	HODAS59	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	103	1544	898	1531	975	975	405	1			21
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	104	871	106	871	248	248	406	1	34	35	174

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	250	865	97	865	258	258	552	1	19	20	177
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	105	404	1	404	16	16	407	1	21	22	64
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	251	2082	852	2074	829	829	553	1	22	23	72
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	106	1542	506	1542	122	122	408	1	51	52	280
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	252	1482	508	1482		633	554	1	15	16	45
96	HCUHB01	209215 08/21/97	ZAP Express	253	834	1	834	82	82	555	1	40	41	251
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	107	2327	1528	2327	465	465	409	1	30	31	284

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	254	1508	885	1508		988	556	1			19
98	HAQBT94	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	108	1062	157	1062	172	172	410	1	28	29	187
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	109	2539	275	2501	903	903	411	1	30	31	237
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	255	2514	592	2431	176	176	557	1	30	31	217
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	256	2357	465	2288		1151	558	1	12	13	82
100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	110	1751	969	1751	4	4	412	1	46	47	192

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	257	689	218	655	314	314	559	1	18	19	95
100	HEONN58	209119 06/12/97	pSportI	258	2377	5	2377	25	25	560	1	28	29	54
101	HCRAM28	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	111	1117	1	1117		1	413	1	19	20	21
101	HIBEK16	209627 02/12/98	Other	259	1193	69	1135	242	242	561	1	24	25	108
102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	112	1313	128	1313	271	271	414	1	30	31	51
102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	260	1262	26	1262	35	35	562	1	35	36	50
103	HEBDJ82	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	113	1654	553	1654	709	709	415	1			32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	114	1171	540	1171	337	337	416	1	30	31	163
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	261	1179	626	1161	335	335	563	1	30	31	253
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	262	1162	629	1131	942	942	564	1			18
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	115	842	373	800	100	100	417	1	65	66	174
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	263	735	290	735			565	1			
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	264	783	416	783		413	566	1	33	34	73

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	116	1640	187	1470	581	581	418	1	30	31	50
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	265	1638	301	1405	119	119	567	1	30	31	263
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	266	1455	148	1188	438	438	568	1	24	25	70
107	HE6DK18	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	117	952	418	906	499	499	419	1	28	29	120
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	118	1256	21	1079	301	301	420	1	30	31	159
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	267	1086	25	1050	227	227	569	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	119	1143	171	1051	175	175	421	1	50	51	154
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	268	1003	21	1003	115	115	570	1	34	35	104
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	269	1234	174	1015	232	232	571	1	27	28	132
110	HSXBL78	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	120	1782	1	1720	138	138	422	1	32	33	204
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	121	610	18	609	50	50	423	1	30	31	67
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	270	574	1	566	337	337	572	1	27	28	32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
112	HOEAP41	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	122	526	185	375	143	143	424	1	21	22	25
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	123	2081	1179	1976	48	48	425	1	30	31	299
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	271	1731	889	1626	886	886	573	1	18	19	28
114	HTXGS75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	124	1717	764	1640	76	76	426	1			13
115	HOVBA03	97902 02/26/97 209048 05/15/97	pSport1	125	804	1	804	145	145	427	1	15	16	198
115	HOVBA03	97902 02/26/97 209048 05/15/97	pSport1	272	1320	77	637	280	280	574	1	22	23	40

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	126	431	1	431	73	73	428	1	38	39	47
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	273	515	1	515	43	43	575	1	20	21	30
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	127	3752	3465	3752	748	748	429	1	30	31	370
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	274	2995	2738	2995	2777	2777	576	1	18	19	29
118	HASAS24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	128	1144	669	1144	896	896	430	1			30
119	HSIDN55	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	129	1830	1234	1830	1265	1265	431	1			24

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
120	HGBGZ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	130	1864	1505	1741	1578	1578	432	1	37	38	53
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	131	2041	1	1214	46	46	433	1	35	36	176
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	275	1990	8	1128	71	71	577	1	16	17	92
122	HOECP43	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	132	2012	853	1986	1127	1127	434	1	22	23	77
123	H2CBV31	97902 02/26/97 209048 05/15/97	pBluescript SK-	133	1669	670	1632	962	962	435	1	25	26	32
124	HPCAD23	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	134	1565	281	1565	274	274	436	1	25	26	30

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
125	HSPAG15	97902 02/26/97 209048 05/15/97	pSport1	135	2007	1101	2007	1124	1124	437	1	39	40	69
126	HELGH31	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	136	1291	1	1180	107	107	438	1			19
127	HUSHH48	97902 02/26/97 209048 05/15/97	Lambda ZAP II	137	1906	1	1906	184	184	439	1	30	31	43
127	HUSHH48	97902 02/26/97 209048 05/15/97	Lambda ZAP II	276	2436	572	2436	726	726	578	1	30	31	42
128	HLYA095	97902 02/26/97 209048 05/15/97	pSport1	138	1935	1044	1794	1183	1183	440	1	18	19	33
129	HHSCV65	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	139	1446	572	1347	585	585	441	1	25	26	53

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
130	HTTAD57	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	140	1109	639	1109	676	676	442	1	24	25	64
131	HEBGA37	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	141	497	9	497	95	95	443	1			34
132	HEBFU93	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	142	269	1	269	1	1	444	1	30	31	89
132	HEBFU93	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	277	782	408	781		571	579	1	31	32	70
133	HSGSC60	97902 02/26/97 209048 05/15/97	Lambda ZAP II	143	1269	55	1262	55	55	445	1	25	26	350
134	HPMGD24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	144	1944	97	1871	306	306	446	1	16	17	49

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	145	1021	526	1021	74	74	447	1	30	31	278
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	278	961	524	961	545	545	580	1	23	24	110
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	146	1285	5	1285	116	116	448	1	30	31	199
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	279	1228	9	1228	324	324	581	1	26	27	30
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	147	1386	169	1272	165	165	449	1	30	31	258
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	280	1327	169	1208	160	160	582	1	23	24	71

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
138	HMWGI25	97902 02/26/97 209048 05/15/97	Uni-Zap XR	148	2098	721	2044	784	784	450	1	18	19	87
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	149	1847	1689	1847	241	241	451	1	33	34	315
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	281	799	1	799		243	583	1	12	13	47
140	HMSKE75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	150	1569	113	1517	417	417	452	1	21	22	52
141	HCMSSH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	151	1540	538	1540	48	48	453	1	30	31	383
141	HCMSSH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	282	2196	270	2196	294	294	584	1	32	33	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
142	HTWCB92	97902 02/26/97 209048 05/15/97	pSport1	152	1719	690	1575	6	6	454	1	52	53	186
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	153	863	1	863	195	195	455	1	26	27	163
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	283	1185	277	1166	621	621	585	1			19
144	HFAMG13	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	154	1101	1	512	40	40	456	1	21	22	46
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	155	2031	669	2031	411	411	457	1	23	24	105
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	284	1634	615	1485	878	878	586	1	20	21	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	156	1981	1458	1809	1592	1592	458	1	23	24	70
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	285	1795	1458	1749	1562	1562	587	1	33	34	69
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	157	915	45	912	22	22	459	1	22	23	155
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	286	858	46	858	224	224	588	1	30	31	77
147	HSJAP03	209139 07/03/97	Uni-ZAP XR	287	915	1	915	22	22	589	1	22	23	155
148	HSKGO26	97903 02/26/97 209049 05/15/97	pBluescript	158	2117	51	1422	32	32	460	1	23	24	332
149	HCQAV96	97903 02/26/97 209049 05/15/97	Lambda ZAP II	159	2395	1509	2382	1440	1440	461	1			5

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
150	HSNCC16	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	160	2120	1223	2108	1416	1416	462	1			14
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	161	900	482	900	46	46	463	1	30	31	285
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	288	1517	783	1517	1062	1062	590	1			24
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	162	1003	1	1003	288	288	464	1	30	31	80
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	289	3865	217	1195	281	281	591	1	16	17	38
153	HTSFQ12	97903 02/26/97 209049 05/15/97	pBluescript	163	2196	1607	2180	1611	1611	465	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	164	1945	271	1840	299	299	466	1	63	64	96
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	290	1910	279	1818	355	355	592	1	39	40	69
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	165	2933	489	2871	258	258	467	1	30	31	399
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	291	3276	486	2838		525	593	1	45	46	308
156	HJPCJ76	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	166	2243	343	2221		341	468	1			1
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	167	1816	1130	1816	284	284	469	1	31	32	273

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	292	1695	1098	1548	1306	1306	594	1			22
158	HMKBA64	97903 02/26/97 209049 05/15/97	pSport1	168	945	1	787	208	208	470	1	18	19	192
159	HNFIP24	97903 02/26/97 209049 05/15/97	pBluescript	169	902	46	816	19	19	471	1	26	27	234
160	HCELB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	170	1883	798	1869	1001	1001	472	1	45	46	105
160	HCELB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	293	1501	438	1501	510	510	595	1			24
161	HAWBA28	97903 02/26/97 209049 05/15/97	pBluescript SK-	171	2100	1642	2100	1722	1722	473	1	23	24	32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	172	1930	187	1930	65	65	474	1	30	31	571
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	294	2683	183	2683	431	431	596	1			24
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	173	1509	962	1451	122	122	475	1	30	31	312
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	295	1454	961	1420	976	976	597	1			1
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	174	3173	2197	2972	51	51	476	1	21	22	329
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	296	828	52	828	305	305	598	1			8

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	175	991	374	970	60	60	477	1	24	25	178
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	297	2416	1387	2413	1473	1473	599	1	18	19	25
166	HFKFX55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	176	1290	499	1290		688	478	1	25	26	52
167	H2LAO11	97903 02/26/97 209049 05/15/97	pBluescript SK-	177	2290	1	2290	173	173	479	1	22	23	62
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	178	549	1	549	11	11	480	1	21	22	27
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	298	545	1	545	17	17	600	1	21	22	27

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	179	1509	294	1352	92	92	481	1	30	31	339
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	299	1530	385	1530	562	562	601	1	23	24	61
170	HCFAE79	97904 02/26/97 209050 05/15/97	pSport1	180	1316	985	1250	995	995	482	1	26	27	32
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	181	777	1	777	51	51	483	1	30	31	48
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	300	997	244	997	300	300	602	1	23	24	29
172	HODCW06	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	182	791	1	791	14	14	484	1	29	30	38

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
173	HFTAR26	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	183	1405	346	1405	575	575	485	1	20	21	61
174	H2MBF44	97904 02/26/97 209050 05/15/97	pBluescript SK-	184	1596	75	1596	131	131	486	1	24	25	346
174	H2MBF44	97904 02/26/97 209050 05/15/97	pBluescript SK-	301	2345	75	2345	233	233	603	1	56	57	69
175	HE8BI92	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	185	2293	355	2288	67	67	487	1	30	31	237
175	HE8BI92	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	302	2369	2	1946		60	604	1	9	10	24
176	HFTBR48	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	186	1212	462	1180	257	257	488	1	30	31	200

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
176	HFTBR48	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	303	1181	424	1149	663	663	605	1	23	24	35
177	HE9CM64	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	187	1605	770	1554	166	166	489	1	30	31	351
177	HE9CM64	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	304	1537	719	1515		787	606	1	43	44	130
178	HATAV51	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	188	1516	960	1516	8	8	490	1	30	31	265
178	HATAV51	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	305	1493	1	1261	54	54	607	1	18	19	23
179	HAQAF27	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	189	681	287	681		401	491	1			25

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	190	1014	703	1014	360	360	492	1	30	31	159
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	306	577	1	577		175	608	1			6
181	HAFAU18	97904 02/26/97 209050 05/15/97	pBluescript SK-	191	2779	2207	2630	1153	1153	493	1	30	31	279
181	HAFAU18	97904 02/26/97 209050 05/15/97	pBluescript SK-	307	2860	163	2860	21	21	609	1	30	31	232
181	HAFAU18	97904 02/26/97 209050 05/15/97	pBluescript SK-	308	876	275	876	302	302	610	1	32	33	34
182	HETBY74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	192	1923	30	1923	45	45	494	1	33	34	193

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	193	2346	1160	2286	178	178	495	1	30	31	205
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	309	2025	840	2025	971	971	611	1	18	19	21
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	194	3054	2004	3054	434	434	496	1	11	12	147
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	310	3026	1966	3026		2131	612	1			9
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	195	907	152	907	297	297	497	1	30	31	64
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	311	712	67	712	107	107	613	1	18	19	29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
186	HFAMH74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	196	1290	84	809	225	225	498	1	30	31	94
186	HFAMH74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	312	1289	785	1289	927	927	614	1	28	29	30

Table 1 summarizes the information corresponding to each "Gene No."

described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinjé, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinjé, *supra*.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinjé are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO: Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

"Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: *COMPUTATIONAL MOLECULAR BIOLOGY*, Lesk, A.M., ed., Oxford University Press, New York, (1988); *BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS*, Smith, D.W., ed., Academic Press, New York, (1993); *COMPUTER ANALYSIS OF SEQUENCE DATA, PART I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, (1994); *SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY*, von Heinjé, G., Academic Press, (1987); and *SEQUENCE ANALYSIS PRIMER*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, (1991).) While there exists a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans. (Carillo, H., and Lipton, D., *SIAM J Applied Math* 48:1073 (1988).) Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in "Guide to Huge Computers," Martin J. Bishop, ed., Academic Press, San Diego, (1994), and Carillo, H., and Lipton, D., *SIAM J Applied Math* 48:1073 (1988).

Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux, J., et al., *Nucleic Acids Research* (1984) 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215:403 (1990), Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711 (using the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981)).

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set so that the percentage of identity is calculated over the full length of the reference polynucleotide and that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brulag et al. (Comp. App. Biosci. 6:237-245 (1990)). The term "sequence" includes nucleotide and amino acid sequences. In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter.

As an illustration, a polynucleotide having a nucleotide sequence of at least 95% "identity" to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone, means that the polynucleotide is identical to a sequence contained in SEQ ID NO:X or the cDNA except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the total length (not just within a given 100 nucleotide stretch). In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to SEQ ID NO:X or the deposited clone, up to 5% of the nucleotides in the sequence contained in SEQ ID NO:X or the cDNA can be deleted, inserted, or substituted with other nucleotides. These changes may occur anywhere throughout the polynucleotide.

Further embodiments of the present invention include polynucleotides having at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity

will encode a polypeptide identical to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

5 Using known methods of protein engineering and recombinant DNA

technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., *J. Biotechnology* 7:199-216 (1988).)

10 Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem.* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

25 Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

35 Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make

phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

5 The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

10 The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, *Science* 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

15 As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln; replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

25 Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, and 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, and 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about"

includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geyßen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., *Cell* 37:767-778 (1984); Sutcliffe, J. G. et al., *Science* 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., *supra*; Wilson et al., *supra*; Chow, M. et al., *Proc. Natl. Acad. Sci. USA* 82:910-914; and Bittle, F. J. et al., *J. Gen. Virol.* 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library.

Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Trautnecker et al., *Nature* 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., *J. Biochem.* 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol. Chem.* 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In

preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.

Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).) Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the claimed invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda P_L promoter, the *E. coli* lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, Streptomyces and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera Sf9* cells; animal cells such as CHO, COS,

293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence *in situ* hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage

analysis establishes coinherence between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinherence is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the

present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (¹²⁵I, ¹²¹I), carbon (¹⁴C), sulfur (³⁵S), tritium (³H), indium (¹¹²In), and technetium (^{99m}Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hydridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, ¹³¹I, ¹¹²In, ^{99m}Tc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20

millicuries of ^{99m}Tc . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments," (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention could be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules

may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, DiGeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from

inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, anaphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1).

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Bimaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., *Corynebacterium*, *Mycobacterium*, *Nocardia*), *Aspergillus*, *Bacillaceae* (e.g., *Anthrax*, *Clostridium*), *Bacteroidaceae*, *Blastomycosis*, *Bordetella*, *Borrelia*, *Brucellosis*, *Candidiasis*, *Campylobacter*, *Coccidioidomycosis*, *Cryptococcosis*, *Dermatocycoses*, *Enterobacteriaceae* (*Klebsiella*, *Salmonella*, *Serratia*, *Yersinia*), *Erysipelothrix*, *Helicobacter*, *Legionellosis*, *Leptospirosis*, *Listeria*, *Mycoplasmales*, *Neisseriaceae* (e.g., *Acinetobacter*, *Gonorrhea*, *Menigococcal*), *Pasteurellaceae* (*Actinobacillus*, *Haemophilus*, *Pasteurella*), *Pseudomonas*, *Rickettsiaceae*, *Chlamydiaceae*, *Syphilis*, and *Staphylococcal*. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat

disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2)/Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The

antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, cardiac rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining

whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the

amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an

amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least

90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1 and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated

polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

- 10 Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

Vector Used to Construct Library	Corresponding Deposited Plasmid
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lambd BA	plasmid BA
pSport1	pSport1
pCMVSPORT 2.0	pCMVSPORT 2.0
pCMVSPORT 3.0	pCMVSPORT 3.0
pCR ² .1	pCR ² .1

- 25 Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128,256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS.
- 30 The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation

of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

- Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lalfmid BA (Benito Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR².1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clontz, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

- The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

- Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.

- The oligonucleotide is labeled, for instance, with ³²P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection

agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μ l of reaction mixture with 0.5 μ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to

remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic PI library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P^{32} using the rediprimer™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of

conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

10 A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

20 The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

25 Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.₆₀₀) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

35 Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic

agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

5 Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

10 The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After

15 renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

20 In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number XXXXXX.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a TS phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

25 DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a

stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Comassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as

required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ³⁵S-methionine and 5 µCi ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV1 and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden).

5 pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QG1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

10 The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., BioTechnology 10:169-175 (1992).) Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

35 A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the

naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

5 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418.

After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Trautnecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the

35 polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having

more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No.209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGGCCAAATCTTCTGACAAAACACATGCCACCGTGCC
CAGCACCTGAATTCGAGGGTGACCGTCAGTCTTCTCTTCCCCCAAAACC
CAAGGACACCCCTCATGATCTCCCGACTCTCTGAGGTACATGGTGGTGGT
GGACGTAGCCACGAAGACCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGGGGAGGAGCAGTACAAC
AGCACGTACCGTGTGTGTCAGCGTCTCACCGTCTGACCCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAGCCCTCCCAACCCCC
ATCGAGAAACCATCTCCAAAGCCAAAGGGCAGCCCGGAGAACCAAGGT
GTACACCCCTGCCCATCCCGGATGAGCTGACCAAGAACAGGTGAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGA
GAGCAATGGGACGCCGGAGAACTACAAGACACCGCTCCCGTGGTGGG
ACTCCGACGGTCTCTTCTCTACAGCAAGCTCACCGTGGAGACAGCA
GGTGGCAGCAGGGGAACGTCTTCTCATGTCTCGGTGATGCGATGAGGCTGCG
ACAACCACTACACGACAGAGCCCTCTCCCTGTCTCCGGTAAATGAGTGC
GACGGCCGCACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 µg/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody

whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab)₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab)₂ fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhitaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10⁴ cells/well in 5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhitaker))/10% heat inactivated FBS(14-503F Biowhitaker)/1x Penstrep(17-602E Biowhitaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression

vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (see below) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

HGS-CHO-5 medium formulation:**Inorganic Salts**

CaCl ₂ (anhyd)	116.6 mg/L
CuSO ₄ ·5H ₂ O	0.00130
Fe(NO ₃) ₃ ·9H ₂ O	0.050
FeSO ₄ ·7H ₂ O	0.417
KCl	311.80
MgCl ₂	28.64
MgSO ₄	48.84
NaCl	6995.50
NaHCO ₃	2400.0
NaH ₂ PO ₄ ·H ₂ O	62.50
Na ₂ HPO ₄	71.02
ZnSO ₄ ·7H ₂ O	4320

5 Lipids

Arachidonic Acid	.002 mg/L
Cholesterol	1.022
DL-alpha-Tocopherol-Acetate	.070
Linoleic Acid	0.0520
Linolenic Acid	0.010
Myristic Acid	0.010
Oleic Acid	0.010
Palmitic Acid	0.010
Palmitic Acid	0.010
Pluronic F-68	100
Stearic Acid	0.010
Tween 80	2.20

Carbon Source

D-Glucose	4551 mg/L
-----------	-----------

Amino Acids

L-Alanine	130.85 mg/ml
L-Arginine-HCL	147.50
L-Asparagine-H ₂ O	7.50
L-Aspartic Acid	6.65
L-Cystine-2HCL-H ₂ O	29.56
L-Cystine-2HCL	31.29
L-Glutamic Acid	7.35
L-Glutamine	365.0
Glycine	18.75
L-Histidine-HCL-	52.48

H ₂ O	
L-Isoleucine	106.97
L-Leucine	111.45
L-Lysine HCL	163.75
L-Methionine	32.34
L-Phenylalanine	68.48
L-Proline	40.0
L-Serine	26.25
L-Threonine	101.05
L-Tryptophan	19.22
L-Tyrosine-2Na-2H ₂ O	91.79
L-Valine	99.65

Vitamins

Biotin	0.0035 mg/L
D-Ca Pantothenate	3.24
Choline Chloride	11.78
Folic Acid	4.65
I-Inositol	15.60
Niacinamide	3.02
Pyridoxal HCL	3.00
Pyridoxine HCL	0.031
Riboflavin	0.319
Thiamine HCL	3.17
Thymidine	0.365
Vitamin B ₁₂	0.680

Other Components

HEPES Buffer	25 mM
Na Hypoxanthine	2.39 mg/L
Lipoic Acid	0.105
Sodium Putrescine-2HCL	0.084
Sodium Pyruvate	55.0
Sodium Selenite	0.0067
Ethanolamine	20mM
Ferric Citrate	0.122
Methyl-B-Cyclodextrin complexed with Linoleic Acid	41.70
Methyl-B-Cyclodextrin complexed with Oleic Acid	33.33
Methyl-B-Cyclodextrin complexed with Retinal Acetate	10

Adjust osmolarity to 327 mOsm

Example 12. Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class 1 cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN- α , IFN- γ , and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	ISRE Ligand	JAKs			STATs	GAS(elements) or
		Tyk2	Jak1	Jak2	Jak3	
5	IFN family IFN- α / β IFN- γ (IRF1>Lys6>IFP) IL-10	+	+	+	1,2,3 1	ISRE GAS
10	gp130 family IL-6 (Pleiotrohic) (IRF1>Lys6>IFP) IL-11 (Pleiotrohic) OnM (Pleiotrohic) LIF (Pleiotrohic) CNTF (Pleiotrohic) G-CSF (Pleiotrohic) IL-12 (Pleiotrohic)	+	+	+	1,3	GAS
15	g-C family IL-2 (lymphocytes) IL-4 (lymph/myeloid) >>Ly6 (IgH) IL-7 (lymphocytes) IL-9 (lymphocytes) IL-13 (lymphocyte) IL-15	-	+	-	1,3,5 6	GAS GAS (IRF1 = IFP)
20	gp140 family IL-3 (myeloid) (IRF1>IFP>>Ly6) IL-5 (myeloid) GM-CSF (myeloid)	-	-	+	5	GAS
25	Growth hormone family GH PRL EPO CAS>IRF1=IFP>>Ly6	-	-	+	5 1,3,5 5	GAS GAS GAS GAS (B-
30	Receptor Tyrosine Kinases EGF	+	+	+	1,3	GAS (IRF1)
35	PDGF CSF-1	+	+	+	1,3 1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994)), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5'-GCGCCTCGAGATTCCCGGAAATCTGCGCATCTCAATTAG-3' (SEQ ID NO:3)
 10 AAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG-3' (SEQ ID NO:4)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5'-GCGGCAAGCTTTTGTCAAAGCCTAGGC-3' (SEQ ID NO:4)

15 PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5'-CTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATG
 20 ATTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
 CTAATCTCCGCAATCCCGCCCTAATCGGCCAGTTCGCCCATCTCCGC
 CCATGGCTGACTAATTTTATTATGACAGGCCGAGGCCGCTCCGC
 CTCTGAGCTATCCAGAAGTAGTGAGGAGGCTTTTGGAGGCTTAGGCTTT
 TGCAAAAGGCTT-3' (SEQ ID NO:5)

25 With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

30 The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a

neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

10 Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

20 Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

30 Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

35 Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI

+ 10% serum with 1% Pen-Strep. Combine 2.5 ml of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at 20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 40°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATs signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM MgCl_2 , and 675 uM CaCl_2 . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting 1×10^6 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 17. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCAGGGATGACAGCGATAGAACCCCG-3' (SEQ ID NO:6)
5' GCGAAGCTTCGCGACTCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40-early promoter sequence, and is flanked with an XhoI site:

5'-GCGGCCCTCGAGGGGACTTCCCGGGGACTTCCGGGGACTTCCCGGGGAC
TTTCCATCTGCCATCTCAATTAG-3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5'-GCGGCAAGCTTTTGCAGAGCCTAGGC-3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2⁺. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

20 5'-CTCGAGGGGACTTCCCGGGGACTTCCCGGGGACTTCCGGGACTTTC
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAATCCGCCCA
TCCGGCCCTAATCCGCCCAAGTTCGCCCAATTCTCCGCCCATGGCTGACT
AATTTTATTATGAGAGGCCGAGGCCGCTCGGCTCTGAGCTATTC
CAGAAAGTAGGAGGCTTTTGGAGGCTAGGCTTTTGCAAAAAGCTT:
25 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-

promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the

NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room

temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4

15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes In Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000-20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2.5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension.

The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mV; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular

signaling even which has resulted in an increase in the intracellular Ca^{++} concentration.

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are

used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of

Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium.

- 5 Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer (20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 40°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 40°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by

- 20 determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 25 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂⁺ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix components gently and preincubate the reaction mix at 300C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- 35 Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 370C for 20 min. This

allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp. (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations In a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products is then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epizentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals is identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, C.V. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect soluble polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on

the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmacologically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric

acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of

about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile.

Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to

transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is being produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

(1) GENERAL INFORMATION:

- (i) APPLICANT: Human Genome Sciences, Inc. et al.
(ii) TITLE OF INVENTION: 186 Human Secreted Proteins
(iii) NUMBER OF SEQUENCES: 644
(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.
(B) STREET: 9410 Key West Avenue
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: USA
(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
(B) COMPUTER: HP Vectra 486/33
(C) OPERATING SYSTEM: MSDOS version 6.2
(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
(B) FILING DATE: March 6, 1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: A. Anders Brookes, Esq.
(B) REGISTRATION NUMBER: 36,373
(C) REFERENCE/DOCKET NUMBER: PS002.PCT

(vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504
(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCGCAATCT TGTGACAAA CTCACACATG CCGACGCTGC CCGACGCTGC 60
AATTCGAGGG TCGACCTCTCA GTCTTCTCT TCGCCGCAAA ACCGAGGAC ACCCTGATGA 120
TCTCCCGGAC TCTCTAGGTC ACATCGCTGG TGTGTGAGCT AAGCCACGAA GACCTCTGAG 180
TCAGTTTCAA CTGTGAGCTG GACGGGCTGG AGCTGCATTA TCCGACGACA AAGCCGCGGG 240
AGGACAGTA CAACGACAG TACCGTGTGG TCAGCGTCTT CACCGTCTCT CACGAGGACT 300
GGCTGATGG CAGGAGTAC AATGTGCGAG TCTCCACGAA AGGCTCTCCA ACCGCTCTGC 360
AGAAACCAT CTCGAAAGCC AAGGGGCGAG CCGGAGAGCC ACAGTGTATC ACCCTGCCCC 420
CATCCCGGGA TCGAGTGGCC AGAGACGAG TCGAGCTGAC CTGCTCTGTC AAGGCTTTCT 480
ATCCAGGGA CATCCCGCTG GAGTGGGAGA GCAATGGGCA GCGGAGGAC AACTACAGCA 540
CCAGGCTTCC CTTCTCTGAC TCCGAGGCT CTTTCTTCT CTACAGGAG CTGACCGCTGG 600
ACAGAGGAG GTGTGAGCAG GGGAGCTCT TCTATGCTC CTTGATGAT GAGGCTCTGC 660
ACAGCACTA CACGAGGAG AGCTCTTCCC TGTCTCCGGG TAAATGATG GAGAGGCGCC 720
GACTCTAGAG GAT 733

(2) INFORMATION FOR SEQ ID NO: 2:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
1 5

- (2) INFORMATION FOR SEQ ID NO: 3:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGCGCTGGAG ATTTCGCGGA AATCTAGATT TCCCGGAAT GATTTCGCG AATGATTTC
60
CCGGAATAT CTGCATCTC AATAG
86

(2) INFORMATION FOR SEQ ID NO: 4:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGCGAAGCT TTTTCGAAG CTTAGGC
27

(2) INFORMATION FOR SEQ ID NO: 5:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATT CCGGAATTC TAGATTCCC GGAATGATT TCCCGGAAT GATTTCGCG
60

AATATCTGC CATTGATTT AGTAGGAC CATATCTCCG CCGCTAACTC CCGCATCCC
120
GCGCTAACT CCGCCAGTT CCGCCCATTC TCCGCCCAT GCGTACTAA TTTTITTTT
180
5 TTATCCAGAG CCGCAGCGCG CTTCCGCTC TAGCTATTC CAGATGACT GAGGAGGCTT
240
TTTTCAGGC CTAGCTTTT GCAAAAGCT T
271

(2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CGCTCCAGG CATACAGCG ATAGACCC GC
32

(2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGGAGCTTC GCGATCCCC GCAATCGCT C
31

(2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGAGCTTC CC
12

(2) INFORMATION FOR SEQ ID NO: 9:

- (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GGGCGCTCGA GGGGACTTTC CCGGGGACTTTC CCGGGGACTTTC CCGGGGACTTTC CCGGGGACTTTC 60

10 CCGTCTCAAT TAG 73

(2) INFORMATION FOR SEQ ID NO: 10:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

25 CTGAGAGGGA CTTTCCGGGG GACTTTCGG GACTTTCGG GACTTTCGG GACTTTCGG GACTTTCGG GACTTTCGG 60
CAATTATGCA GGAACGATAG TCCCGGCGCT AACTCCGCCC ATCCCGGCCC TAACTCCGCC 120
CAGTTCGCCC CACTTTCGG CCGATGGGTC ACTAATTTT TTTATTTATG CAGAGGCGGA 180
GGCGGCGTGG GCGCTTGAGC TATTCGAGAA GTATGAGGAA GCGTTTATTT GAGGCGTTAG 240
CTTTTCGAAA AAGCTT 256

(2) INFORMATION FOR SEQ ID NO: 11:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 582 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

50 GCGACGAGGT AATTCTTACC AGAATTTCC AGAGCATTTAT GTAGTNGAA AAAAAATCGAA 60
GCAAGCTGTT AAGATCTCTG GATCCGTTTA TATAGTACGT ATAGCTGAAA TCTGTATTC 120
AATCACTTTT TCTCTTTTAT CTTCTAACCA AAAAATTTGT TAAATTTTGA TCCCAATGT 180
TTTTTATCTT TGTATATTTT TTAAATATCC TTTTCTCTTC ATCATGCTCT TTTTGTGCT 240
TGTAAATAGA CTTATTTGCA CTTTGAGAT GAGTTACTCC TTGTCAATCT ACAAATATGT 300
GATATGGAAT TTTTCTATAC AGATGTACGT TTTTGACCAA GAAATGTGTA TTTGTTTATA 360
AGAAAAAATC TGGCTTCATT TCTGTGAAT TCGTCTTTCA AATTTCTTTT TTACAGCTGT 420

AAGCAACTG AGATACCGTG ATGGTGTGGA TTCTTTTCA TGAATCTTAC CACTATTTT 480

5 AAGCACTGAG CTTTATTTA TTGTCTAAT TGTAAAGTTT AATTGTCTTA ACTCATTTA 540

TAATATPACT GTTATCTGT TTCTGAAAA AAAAAAAAAA AA 582

(2) INFORMATION FOR SEQ ID NO: 12:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

20 GTTTGGGGGT GAGCGCGAGC TGCCTGGGG CTTCTGCCC GCGCAGGACA CAGCTACTCG 60
CAGCGCGGCG GCGCTGGCT ATGATGTTCC TCAACAGGG GGGGCTCTTG CCTCTACTC 120
GTCCAGGCC CACTTTCGAG GAGAGAGCCC TCCCAAGCC TTCAAGGCTG CTGGAGTCA 180
CCTTTTGAAA TGGACTTAAA GGAACCTTGT GTGGAGACAG GTGCTTCCCA AAGAGCTTC 240
TCTGTGCTTC CAGGAGGCCC CTTCTGAAG GAAAGGGGAG GACTCATCAG GAGCTGCTTG 300
GAGCCCTTCA GCGAGGGGAG CTTGGGCGCG AGCCCAAGCA TTTGCTCTTG CTGCCCCGA 360
GGGACAGGA AGCCTCTTGG GCTCTTCCC TTCTTGACA AGGCGGCGCT CTTTGGCTC 420
35 AATATACCTG TACATATTT TCAITTAAG CCTCTTTGAT AAAAA 465

(2) INFORMATION FOR SEQ ID NO: 13:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

50 ATCGAATTC TGCCTACAGC CTTTCTTTG GTGCCACTTC TGCGTCTTTG TGAATGCCCC 60
ATATGCTAG CTTCTGCCC CTCCTAGAG GCTTCTTGA TGAATTAAGA AATAGAAATG 120
AGTGAATTT CCTATGTGCA TATAGAGG ACCCAAGAG CATCTCTTTT AATTAAGG 180
AAGATGTCCA TCCCTTTTAC TCGGAATAG AACTTGTGTC TCAATCTCTT GAGACTAGGC 240
CTTTAAGCA CTTTCTGTGT TCTCATTTG TCTCATGTTT TTGCCAGGCT TTTATCGGA 300
60 AGATATGTT CCGTTTAAA TATTTCTTA TGAAGCGCGG CGTGTGTGCT CAGCGCTGTA 360

5 ACCTGACAG TTGGGGCTG ACCGGTGA TCACGAGTC AGCAGATCA GACCATCTG 420
GSTMATCG TGAACCCCG TCTCTACTAA AATTACAAA AAAAAAAAA AAA 474

10 (2) INFORMATION FOR SEQ ID NO: 14:

15 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

20 TTAATTTGG GAGCAGACC TGAATGACG CTTTACATG GAGATTAAT TCTGTCTCC 60
ATCTCAATG CCCAGTACC TGAACGAGA TGAATATAC CACACACAT GTCTCTTAT 120
CATGATGCG TTATGAGTA GATTATTTC ATCATTGCA TTTTATAGTC TACATGCTT 180
TATATTAAT TCTCATTTT TACGTCTTC CTTAGTGGC TGTGTATATC AATCTCATG 240
CTCTCAGG CAGTACAT CTGATATC CATTATGCG GAGAGCTTG TTGAGAGAG 300
AGATATATAC TTCC 314

(2) INFORMATION FOR SEQ ID NO: 15:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 613 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45 CTGATATCG CTTCTGGCTA AAGTGAACA TCCCATTAAT CAATCTGCT TTAATTAAT 60
ATGTTCTTAA TGTGGCAG CAGACAGA AATGAGAGA AAGTGGTGT AAGCTCAGA 120
ACCACATAA TCTATCTAT GGGTGGGTT CACCCAGCTT GCTTTTGA TTTTGTCTA 180
CTATACAGA GCTCCAGG AACTGAGA CTCAGCTCC TTATGCTG TAACTTAAGC 240
CTACTCTTC GATTACACC AACATATTC CAGCTATC CTCCTTCC AAGTGGCTT 300
TCTGGCCAG ATCGATATG CATTATCTA ATCTCAATG GAATGTCCA CAGGATTC 360
CAATTTAAT AGCATACAT AATTTTGT CAAATTTCT TTGAGAGAC TCCCATGCA 420
GCTGCTAGA GACGTATG GAGGGCTC TTGAGCTTT CCGATAGT TTAGCTGCA 480
ATAGCTCTTA GCGAATCG CAGAGCTC CTCGCAACT GTATTACTG GGAACACTG 540

ATTGCTTGA AGTTATCTT CCTTAACCT TTCTAGTTC TTACATTTA CTGCTGAAC 600
CCAAATATCG CAC 613

(2) INFORMATION FOR SEQ ID NO: 16:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

20 CCCCCCAT TGAACCTCG GCTGTGAAG TTTTGGCTG TGTGGTCTT TCTGTGTGC 60
GCTGTGTG TGAATGCCA CTCCTTTGC AAGTGGGAG CAGCCATCA TGAAGCCGCC 120
TTATTTTAA TTGAGATGA CCAATCTCC CCCCCAGCG CTCGTCTG TCCCTCATG 180
GTGAGTGT TCCCTGCCA AGAGGCTGA TTGTGGGAA ATGCCATCAT CTATATGAT 240
GGAGGCAAT TTGTCTCTG TTAATTTAT TACACATCA TTGCACTCTG GCACTCCAT 300
CCTGAAAC CTATTTTGT GTTTACCA AGACACAG GGGAAAAA AAAA 356

(2) INFORMATION FOR SEQ ID NO: 17:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 414 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

45 GAATTAAT CCCCCGCTT TTACAGTGA CTTGGAAAT AAGTATGGG TAATCTATA 60
GAGCATTT ACTGACCA ACCAAGCTA TGAAGAAA TGAATTAAT TTAAAAAT 120
ATTACATTA ACTGTCTAG GATTACTCT CTCAGGCTT GGAATCTTC TTCTTGAAA 180
TTTGATAT CACTTCAAT CTGTACAA AATTTTAT CTTGAGATG CAATTTCTC 240
TCTCCAGAA AATGACTA CACAGCTC AACGATAG CTTGTGCTT CAGGGHTTA 300
CACTATGTT TTCTCTCT TCAATGCT ATTTACAGA GACCTTCTA TCAGAGAG 360
TACTGACTA TCTTTATAC TTGTGATTO ATCAGGCTT TAAAAAAA AAA 414

(2) INFORMATION FOR SEQ ID NO: 18:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 469 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

5 ATACCACTT GCATACAAA TGATCTGCTT GGTGAATCTT GAGCTGTACC CCACTTTCTT 60
CAGGAACTTC GAGATGATGCG TGTGTTCTTC CTTGTGTTGAC ATAGATGCGA TAAATACCCC 120
CTTCATAGTC TTACAGCTCA GAGAGTCTG GCAAGCCTTG CCCCCTCATTT TTTTTCGGCT 180
GTTGGGCTCG CTTCGCGCGG GAGTGAGCCT ACTTCTTCA GAGACCAAGG GGTTCGCTTT 240
GCCAGAGACC ATGAAAGGACG CCGAGAACTT TGGGAGAAA GCAAGGCCCA AGAANAACAC 300
GATTACTTCT AGGTGCAAA CTTGAGAAC CTGCGGACAC TGAAGAGATG GTTTTTCGGC 360
GATGTGCTGT TGAAGGAGATG AAGATGAGT TATCTCTCTG AGAATTTCTT AGAGCCTTC 420
25 ACTTCTCTCT ATTCTCTCTC ATACTTCTCT ACCCCCAAT TAAATATGAG 469

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 550 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

40 CCCCCCCCCC CCCCAGACT TTCAGAGTC ACCCCCAAGC ATTTCGGGTT GGTTCGGCC 60
TACTCAGCC TGAAGCTCC TGAAGGAGCC TGCATCTCTT GCTCCCAATC CCCCCTACTG 120
GTCCAGGAT GCAGCTGGA GCTGGGTC TTGTTCTGGG CTTGCTGCTG CCGCCAGCCC 180
AGAGCCCGAG CTTCTCTGA ATTGATCA GTCTTCTCTT GAACTGCTTC CCCCACCTCT 240
GGCATTATTC CCAAGAACT TTAATTTTC TGAAGCTTAA GCACTCTCTG GACTTCAGGG 300
50 ACTGTGTGAG GTAGGCTGAG TGGAGCTCA GTCTTAGAG GTCTCTGAG ATCTGAGACTG 360
AGGACTTCC TACTCCCAA GCGAGAGCC ATCAGCGAGG CTTCTCTGCA GCGACTGCC 420
TGTGAGTCC TGAAGCTAAC CAAAGCTCG CAGCTCTGAG CCCCCTATTA AAGGATCTG 480
55 ATGAGCCGAT GGGCTCTGA GCGAGCCCT TAAAGCATCT GGTCTGTTTT TGGAAAAA 540
AAAAAAAAA 550

60

(2) INFORMATION FOR SEQ ID NO: 20:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 741 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 TCTTGAAGAG TGTACAGTAC AGGATTATTA TATGAAATCT TTATATCAG AGGTTTCTCT 60
TGGCTCTGCA TATATTTATA GCAAAAGAGA TTGTAAATCT GGCACAGTAT TCGAGATAC 120
TTTTCAGTTC CGGCTTTCTT TCTCTTCTTT TAAATTCGAA CTTAGATACA TCGATGAAAA 180
ACTAGAGAAA TGACTTTTAC CTTTGGGAGC AGCCAGTTTT TTTTGTAAAA CTTATTTTCT 240
20 AGCATGCTTT CAGCAAGTTC TCCAGAGCC TAGATTTGCA AGGAGCCACT GTTCTTCTCT 300
TGTACAGCT CCTGAGCA TTGTTGAGAG GACCAATGTC ACATGCTTTC ATGGCCATGG 360
25 NCATGAGAG CACTGCGGTG ATATCTCTCT ACATATTTGG CTTCTCTGCG AGGCTATAC 420
ACAGGCTTC TCTTTCAGAT GATCATTTTC AAAGCTCCCC CAGGCCCCAC CATCCATGCT 480
GAGAGAAA CAGAACTGCT CTGAGAGCA GTCCAGCTA CAGATACACA GCGTTGTGAT 540
TGGGCTGTC ACCTTCTCTC TCCACTCT CTATCTCTGAG AGATGCTGCG TGGATTTTTC 600
CTTAAGCTCA CTTGACTTCC CTTGAAATCT CTATGCTAG TTGAGGGGCT CAGAGCATTC 660
35 ATTGTAGAG TGTATACTCC CATTAGGCT TCTGTTATCA TTTGTTGTGCT TTTTCTCTCT 720
ATTAAAGAA ATGATTTTTC C 741

(2) INFORMATION FOR SEQ ID NO: 21:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 991 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

50 GCGAGAGTTC TCCCTGCGG AATTTTTCT TTTTCAAGAG GAGAGAGGCG TTTCCAGCT 60
AATGTTCTTA GAGTTTGGG CAGAAATCT GCGACAGAC CAGACAGTT CTCTCTTAA 120
TCCAGTCAAT TTTGCTTCTA TCCAGCTAT GTTTCGAGTG TCCCTGCGGT GTTTCAGCA 180
GCAACAGAA ATGATTAAT CTCTGCTGAG TTTGTTATTT GTTCTCTACT TGTGTTTACA 240
60 CTGTATTTTC TGAATTTATG GGTCTCTGCT AATTAAAGAG GAAAGATGAG AATAGTTAAA 300

ACTGAGTTG AAGCAAGAT ACAAATATA GATTAAGCTG ACCTTGAT ATAGCACTT 360
ATAAGCTTA GAGTGTGCTA AGTTGAGTGC AAATTTTCTT CTGATCTTTC TGATGCCAA 420
CAAAAAGCA CTCATGTTTG TTAATGATTT GGAATGCCAC CCGAGAGAG AGCTGCTGT 480
GTCTTTGCG GACAGGAAG CTTCGTGCA CCAAGTCTGA ACACCACTT TGATGTGAC 540
ATAGATATG TGCTGAGCA TATTTCACAC CCGCTTGCA GTAAACACTT GTAGTTGTT 600
CGATGAAA CCGTCACTT CCGTAAGC AGCGGTGTT GTGAGGCA ATGTGATC 660
TGCTGTAAA ACAAAGCTC CATGTATG TGATGCTT ACTGAGAG TGATGTGCA 720
AACAGCTTT GCGAGTCTT CTGTATTTA TGATGAAC CTGGAATC TTGAGACTG 780
AGTTAACAT AGCTCTTAA AATAGCTCC ACAGTTTAT CTATGTTAT CTCTACATC 840
AGGCTGTGCA CGAGCTGTT CAAGTATTA TTTTCTGGA AAATTTTCA GTGTTTAT 900
CGACTTAC CCACTCTG TGAGTTAT TCTTCAAC TCACATTA TGATGATAT 960
AGTCAAATTT AGGGGACTG TATTGCTT A 991

(2) INFORMATION FOR SEQ ID NO: 22:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 653 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CCACCGTCC GGAATTCCTC TGAGATCTT GGGCTATCTT TCACAGGGA TTCTTGCA 60
TTGATGCTTT CTACAAATCA ATATAGTAC TGCCCAAGA TGAGAGCTT GATTTCTAC 120
AGATTGATG AAGATGAGA AACACACAG ATTGAGATA CGGAACCAT GTCTCAATT 180
CTCAATCTTA AATTTGTTCC TCTGAATAT GATATATCC TGATGATCC AGCAGAGAT 240
GCTGAGTAC AACTGATCA GATGATGC AAACAAAGG GAGTGTATG AGACACGCG 300
GATGATTTA GTATTTTAC CAGTGTGTC AAGCGGAGG AAGAAAGGT AGCAGAGAA 360
GTTTGTATG ATCTCATCTG TGATGTGCG AGTCAGGAG TTCTGTACCC AGCTACTCA 420
TCTGAGCAC TTTCTATGTT GTTATATG GAGCAAGCTA TGGAATTTA AGACATCAT 480
CGAGAGAGG GCTTCTGAG GTTCAAGTCC CTGAGACAC TTGTGAAGT 540
CAAGGAGAG AACTCAAGA AGAATATG GAGAGTTTC GTTGCACCT TTCTCTGACT 600
GAACTCAT GTCCAGGCTT GTGCTTCGG AGGCATCCA AAAAAAAAA AAA 653

(2) INFORMATION FOR SEQ ID NO: 23:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGCAGCTGA CCACTGCA GCAATGTC CTGCTCTGCT CTGCTGCA GTTGGGGAC 60
CTTGGGAGG AAGCTGGCTG AGCTCCAGA CCGCGGGGC CATGGCGCG GATCTGCTG 120
TTGGCTTGG GGGTTGAGA CCGCGAAGC GCTTCTGCA CAGAGAGAG TCTCTTCCG 180
GCTGGGCACT GTTCTGCA SGAATGCGA TTGCACTCAT GTTCTGCTAT CCAAGATGC 240
TGTGTTTGG GGGTGTCTG CTGTCTATG CACTGTGCA CTTTTCAGC ACATTTTGC 300
TGATTCCTAT CAGATTCCTG ACGATGGCT ATGTGACCT GTTCCGGGC ACATGTGCG 360
CGAAGATCTT TTGCTGTGCT ACTGAGTCA TGGTGTCTG CTGCAAGCC CTGCTGCTG 420
CGGTGTGCG CCGAAGCTG GATTTTACA AGCGAGAGA CCAAGTCCAC AACTTCATCA 480
TGATATTTCA GTATACAAA GAGTGAAGG AATGCGCTC CCAATGCTCA CAGAAAGCTT 540
GATATTTCTA CAACATACT CCAAGAGAG AATCTCATCT TCCCGCGAG CATGACGCA 600
AATCTGTGCG CCGATCAAC GGTTCGGC AGTGTGGCT GAACACCGG AAGCTCCGG 660
AACAGTGA CTCTATGCT GACATCTCA AGATGACAT GATCTGTAT GACTGTGCG 720
AGATCTGAG CAGCTCAAC CCGCGCTCG AGAAGAGAT TGACAGCTG CCGCGAGCG 780
TGATGCTCT GACTGAGCT CTTAGACTG CCGTGGGCG GAGGAGCTT CCAAGACCA 840
CGCAGCATC CAGTACCTG GACCCAGAG GAGAGACAG GCTACTTTCC CCAATACTCA 900
GTTGTTGAC ATCTCTCTG CCACTCTCA CCGAGCGCTG AACAAAGAC CTCAGTCA 960
AGAACCAAG GGGGCGCTG CTTGAGTGG GTTGGCTTCC TGATGCTTC TGAGGGGAC 1020
GCTGGTAAA GTGGAGGCG TTGGCGCAC CTGAGCGCCC AGTGGGAA ATGCTCACCC 1080
CCACTCTCA TACCTCATC AAAAGACTC TCATATGCT GCTATGAGG ACCTTCAGCT 1140
CTCAATTACA AGTCAAGCG ACTGAGCCA GCACTCTCTG GTCCCTGCA AAGAGGTAC 1200
TAGGGGCGCG GATCAGAGT TCTGGAGCG TTCACTTACC GCTGGCGGAG CTGAAGACT 1260
GGATATAGG CTGGGCGCG GCTGAGCTG CCGCGCGCTG GTGGAGAC AAGAGGACA 1320
CGATTTTTC AGAGCTGAG AGAGACTG GTGGGAGGA AGAAGTGTAA CTCACAGCC 1380
TCTCTCTTA TCTTGTAT AAATTTAA CCAAGAAA AATATAAAA AAAAAAAAA 1440

AACTCGAGGG GGGCCGCGAC CCAATGCC TATATGAC GTANN

1486

5

(2) INFORMATION FOR SEQ ID NO: 24:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2323 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CTTCCCGCTT TCTCTGCA GGGAGTCC CGCTTCCG TGGAGCTCC GACCAAGCC 60
CCTTCAGCTT CTGCTCCG ATCGATGCG TCGCGCGCC GCGCGCGCG TCGCGCTCC 120
TTCCGCTCT GCTCCCGCA CCGCGCTCG CCGAGCAG CAGCATGCG GGGATGAGA 180
AGCAAAAGAC GAGAGCAG CAGAAATCCA CCAATGAGT CTATGAGCC CACCATGGA 240
GGAGAAATTA GAGAGGCA GTGATGAA CAGCGGTG GTTCCGAG TATACCTGT 300
GGTAACAGG TCTCTGCT GCTGGAAA ACAAGATA GTTTGCGCT GAGAGATAC 360
TTTCTTCCA TGCATCTT GTAAATCT GAGTGGGAC AATGTCCTC ATGGCTTAA 420
CAGATGCCC CAGATGCTT CATGCGCCC AAGCATGA AGTCTGAC AGATATTAC 480
AGTCTCTGC AGAGACTA AGCTTTGT CCAGATTC TTCTGAGT GCTCTTTGAT 540
TACCTTTCT ATTATTGA TTGATGCTT TGTATTAT TCTTCTCA TGTGCTTT 600
TATCTTTTA TATGAGTA AAGTTGTCT CTATATTAA AAAATATTAT ATATTATAC 660
ACACACAT ATACATACA AGTAAACTG AAGACCAAT CTGACAGT AAAGCANTA 720
TTCTTACA TTCTATATA AATGAGCTC TATGTATT TTCTGAGC TGAAGGCA 780
GCTCCAGAT TTCTTAGGC TTGTTTGG CATGTGCTA GTTACTCT GAAAGTGA 840
TATATTTC AGCATGCTT GAAACCTGT ACTTTCOA TGTATTAT CAGTTGTA 900
ATATATGCA TGTCTCTCT TCTGTACAT TGACAAAA AAAATTTTT TTTTCTACT 960
CTAAAGAG TGTGCTAC ATGAAATTC TTCTGATAT TTGACTCT GCTGTACAA 1020
GCTTATGT TATATATA TCTTACGCT TGAAGCTG ACTGAGAA CAAATGTGC 1080
AATAGCTGA ATTATTCT AGAGATCTGT GCAGCTATT TCTGTACA AATATTAT 1140
GTCTATTAG AGATGCTTA ATGCTCTG TGAATATCT AACTGATTT ACGGTGAC 1200
TTTTATTAG ATAGATAT TTGATGAG GAGCTCAAC AATGTGCA TGTCTGGA 1260
AGTATCTTT CCGCTCTTT CGTGTGCA TTGTCTTG CAGAAAGAT GCGCTGATG 1320

CAGCAGCC GCGCTCTA ATAAAAAT ATTGACTA TCGAGTAC AGGCATAG 1380
AACTGAAA GACGAGAA AACAAAGAT CCAAGCTTT CATCTTACG GTGACAAAC 1440
TGTGATAG CACATATG TTTTGTAA GCTGTGCA CCGTACAA ATGTAAATT 1500
GCAATTAT GGAAGTCT GTGAGTGA AGAGCACC AGGCATCT ACTCCAGTC 1560
TGTGCTCG TCTACACAG AACACAGG AGCTGGTCA GATTCCTCC AGCTGCTTA 1620
CAAAATCT CAGACAAA GTCTTACA AGCTGCTTC TGGATACT AAAGTCCAG 1680
TTTTGTAC TGCATGAT TTATGCTGT TGAAMAAA AAAAGCTAT TCGAAAGAT 1740
TGAAGCTTT CTGAGATC TTCTATGC TTACTTCT GAGAGCAAT GTTTTACT 1800
TATGCTAT TCAATTTCC CAGGATTA AATGAGAA CAGACCTTT TATATATAG 1860
GTCTCTCG AAGAGACTA AATTAGAG AGAAACTGT GCAATTTTC ATATTCTAT 1920
TCTAAAAA CATTATCTT ACTTACAA AGTCTTTTG AGATTAAGT ACACAAATG 1980
GCGAGAG TTCTCTTTA ATGATATAT GCTTACTC ATGTATCG TTACTACTA 2040
CTGCTTTTA AAAAAGAC CAGATATT ATTGAAACA TGAAGAGA TTATAGTCC 2100
TTAGCAATA TATTTTCTA CTAAAAAT ACATTTAAA GCTCTTCT GCTATATAC 2160
CATGCTAGT GCAATGAT ATTCTATGT ACAGCTATG CTGTTCTTA TTTTATATA 2220
TTTATGAG TGAAMAAA AAAAAGAAA AAAAAGAAA AAAAAGAAA 2280
AAAAAGAAA AAAAAGAAA AAAAAGAAA AAAAAGAAA AAA 2323

(2) INFORMATION FOR SEQ ID NO: 25:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GCGCAGCC TGTGTGCA TTTTCTCT GTCAGATAC CTGACATAG CAGCAGCC 60
TCAATGCTG ACAGCATTC CCGAGCTG CAGCTGTGT TGTGTGTGA AGAGAGGG 120
GGCCGAGG CCGCTTTTG AATGTTTGC CTGTCTGAC TGTGAGCA CTTCGAGTG 180
ATTGTGCT AATTTCAG CTGCTCTGT TTCTGTACA TCTTGAGG GAGCTATGCT 240
CAGACATGC GCGTCTTA GAATGAAA AGTCCCGGT CTGTCTCTT CACTGTGCT 300
CTCTGCGG AGGAAAGAA TGGCTTTGT GCTTTTGT ACAGAGCTA TCGTGTGCT 360
GAAGTGTCC ACATGAGCC TGTGTGAG ACTGTGACA CGGTTCAGC TTGTCACT 420

5 CAGGCTTTC TGTCTCTGAT AGGCTGGAGC AANAAGTGGAA AGCAAGGAA AGAGGCTTTT 480
CTGACAGCA TTATATATAA TATAGAGTGC ATTACATCT GTGCTCTCG GGCACCTTC 540
CCTGTGCTTC AGTCATCTT AGATGACTGA CTCGCAATAC TTGTACAT TTCTCTGGAG 600
CAGCTACTA GGGGAAACAA GATGTATGTC TATTTCCGAT AACAGTAG ATTTCACAA 660
10 CTAAAGAAA AAAAAAAAAA AAA 683

15 (2) INFORMATION FOR SEQ ID NO: 26:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

25 CTGAGAAAG AAGCAATGC GATCTGCTGC AAAAAACAT ATATCCATA AGACTCATGT 60
TATTCAGAAA ACGATATGTA ACGACATCA CATTCGATG ATCTTTTAA AAGGAGGAG 120
ACCTTAAGT ATCTCAAT CTGAATTTCT ATTATTTCT TCACTGATA TAGAACATAT 180
30 GGTATCTGA TTATATAGA TATATTTTG GATATGTTAC TATTTACTT CTTATGCTTG 240
GTAAACATCA TAAGTCTGT TATTAATAC AACATATTC TTTTITAAA GAGAAAGC 300
35 TTATTTTTC TTGCAATGT ATGATTTAT CTACTTATG GTTTTCTCT ATTACTGTTT 360
TAATTTTTTT TTTAAGTTGA GTTTTTCATA AATTTTAGA CCTGTGCTC ACCTTTTITT 420
GAGTCTCTTG TTGCTACTAG GTATATAGCT CATTTTAAA ATCTTAAGC AAAAAATTT 480
40 TATTTATAA AGAATCAAC TTTTGCATGC ATGAGCTGT GAGTCAAT ATTATGAT 540
AAAAAGCA GTCTCTTTT TTATTTTAC CATTGACC CCACCAATG CACTGTTTT 600
45 ATATTAAGA AATATTAACA ATTATTAAT CTCAGATTA AATCTATTC ACTACATCT 660
TTTCCCTCTT TGTCTGATTT TAAGCATGT GTACTTGGCA TCTTCAAT TTCTTAAGCA 720
50 CAGTGTGTTT CTACATATTT AGCATATPAT ATTTTATTT GTCTCTTTT ATTCTATG 780
GCTTCTTACC AGAACATTA GGAAGAAC CAATGCTGT GTATAGACA TGAACATTTG 840
CTCTCATAT GTTTTTTTT CACATCTTT TGAATTTCA CTTTTTAAC GAGAGCAGC 900
55 AAGCAAAATA GATGTGCTTG GTTCTGCTTG TCCGGGGGG TTTTTCACC GAGCTCTAA 960
ATCTCTGTA TTGAGGTTTC CTTTTTGTA CTCAGGATG GAGCTACAG TGGGGCCCC 1020
60 TCTCTCCAT TCTTTTGGAG AGCACTGAG GAAACAGG GTTCTTTTG AGGTGTCTTT 1080

GGCTGCTTT TACGGGATGC GAGCTTCTC CGATCTTTT GTTCTCTGC ACCTCTTGA 1160
GCTACTGCG GTGCAAGTT GTAGATTTTA TTCCGAGGA CCTGCGCTK GGGGCTGAG 1200
5 CTGGGCTAAA TCGAAAGCA TCGAACAGA AGCGGGGCA GGGAGGAAA AGCAGGCTG 1260
GCTCATGCG TCCCTGGAG ATGTCTGAG CATCAAGTC CAGCTGGGC CTGGGAGGC 1320
10 ACCCTAGCA AGGGGCTCAG GTTGCTCTT TACAGAGA ACCTCGAAA CGATATTTG 1380
CACATGAGC TGTGATACA CTATGTTGA TTTTTTTTT TTTTACAGT CATCGAGAT 1440
GTTTCMAAG TGAATTTTAT TTTTTTTA TTCTTTATC TTTACTTAA GTGATATG 1500
15 TATTCCTCG GAGGATAG GAGAAACA GGAATTTAA TATGTCAAA CAGAAACTT 1560
CCTCTTAT TATATATA TCTCATGTA TTTATGCTT AATGTAGCT GACTTTTAA 1620
AAGCTTCTT TTGTGCATG CCTGTGAG CGATCTGAT TGTACATCA TCCCTTTCT 1680
20 CCTGTTTCC TGTAAAGT TACTGAAA AGAATATTT TTGCTGAT TCACTGTCC 1740
AAGCAATCA TATTTTTTA ATTGTCATA TATGGAAGA GCAATTTGT TACATGTAA 1800
25 ACCTTACTG ATATACAGT ATACTAAT TTGAAGATC TGTCTTTTC AAGTGTACG 1860
TTTTCAATG TTGTACAG TGAACACC TTGTGTTTA AACTCTTAC AATGTATTA 1920
TATCTATTT CTTCCATGT AACTAAGAT CATGCTATA TTTCAATCA AGCTATATT 1980
30 GAAAGTAG GCAATGAT ATACAGAT TTTTATACA AAAAAAANA ANUAA 2036

35 (2) INFORMATION FOR SEQ ID NO: 27:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

45 GCGAGGAT AACTAAGCA CATATACT GTATGCTAC TTCTAGATT ATAGAGATT 60
AAGTTGAGA TCAATTTCC ATTGAGAG AATATGTTG CTTTCACTC CTTTATTTG 120
50 ATTCTGGAG AGAGCAAGT CCACCAACA TTCAACCCA GCGCTGAT GACAGTATC 180
CTGAGAGCA GAGCCAGCA GAAACAGCA ATGTAGAAA GTTACAAATG GAAGTTTCC 240
TCCCACTTC GGAATGACA CTCMAAGT GATGCGAAA ACTGCGAGG TATTTCTCT 300
55 CATTAAGT CTAGCCACC ACTTCACT CCGCAATTA ACTAGTCCAG TTGACTAATC 360
CTCTTACCT TTATCATTTA GGTGAGCAT TGCAGAAA CTCTGACTT TCCCATTA 420
60 GGGCTGTGTT TCTCTGTGTT CTTGATAG AGGCATGAC ATATCTGGA AACATGAGT 480

AAATGACGAT TCTTATCTTT CTTCCGACAG CTCCTGAGTT AGAATATCAC AATGTTCTCA 540
GGTAGTCTCA TACATGCTTA AGTTTGAGAA CCATGAGTA AAGTAAATGC ATTAGAGCA 600
GATTAGATAG GAGGTGGCC GTATCTTCTT ACAGTTTCCC TATTACAGG AAGTCAGAG 660
GTCAATTCAT CAGACATAG ATTATTTATT GCTAAAGCTA AAAAAATTA AAAAAA 717

(2) INFORMATION FOR SEQ ID NO: 28:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 495 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTCGGCA GAGACACAT CTAATTTTA GTTTGGAGAT GCATCTTAA GATCTCTTC 60
TATTCCTTTT TCTCCACAA TTATCTTCA TTCTGCTCTT CTGTGACAT TTGCTAGAG 120
AATCACTGCT TCTTTTTCAT AGTAATATCA GAGACTGAT TTTTTCATTT CTCAGAGAA 180
AGGCATTTG CTCCTACAG CTCAGAGCA CCCCCTGGTG GCTGGGGCC TGTGGGAGT 240
TTCTGGGGGA TTGACCTTTA CACATGCG TGCCCTTACA GAAAAACTG CACTAATAA 300
TTATTTTATA AATGGCTCC TCCAGGAAAT GCATATAGG GCTATATACC CAGTATTTTG 360
AGCTTTGAA GAGTATATAT ATCCTGAG AGAGAACTG AGCATATTA GAGGGTGGG 420
ATGACTCATG GTTGGACAC TATGGAGTC TCCCGACAG TAGCAGAT CACCTCACTA 480
CACAGAAATC CTAG 495

(2) INFORMATION FOR SEQ ID NO: 29:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 556 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AGCTTACGCT CATGATATAT TAGGGGAGTG CAGGGCAAAA CAGATGATAG AATGCCCCA 60
GACACCTCTT AGAAGAGCTG CTAGAAAGGC AGACAGCACC AAGCCCTTAA ATGAGATGG 120
GGCATCTGGT CTTCTCTCTT GCTACTGCTT AGGGGTGGAG CAGAGTGGTT CAGTCTGGG 180
CAGTTAGCTG CACATCAGCT GGACCCACCA CACCATTTT CTGGTACTT TACCAAGAG 240

ATAGAAAGC AGCCAGATCT TTACAGCAGC TCTTACTGCT TTCCAAACA ATGCAGATCC 300
CCACATGTC ACAAAGATG ATGTGTCTGT CTTGTGCTAT GAGCAGATG GTGGCTGAGC 360
GTCAAGATC CCACTCTCA AAGAGGCGAG CAGATACAG GCTGCACT GTGTGATTC 420
ACACATGCA CATTCTGAC AGCAGATGC TGCATGGCAA AACAGCATC GGCTGAGAG 480
GACTGCTGAG AAGGGGAGCC GGCTGCTGCG GATGTGGTTT GATTTAGCA GTAGCTCATG 540
GAGATGTGAC CTCAA 556

(2) INFORMATION FOR SEQ ID NO: 30:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTAATGCTG ACTGTGCTT TGTGACACA GGGCCCAAT GTTAGGTGTG AACACAGAT 60
GCACAGATG AGCAGCATG CAGATGCTG AATTAATGTC CTGACAGAT GTGTACATG 120
ACTTCTTTT CAGCTTATT CTGTGGCTG CCTTTGAGA TAGAGCTTTG TTGATATTA 180
CATTAAACA AATTGTATTA TTATGTTTCA TTCTGCAATG TTATTACCA AAGCAAAAT 240
GATTTATTTT ACATGAGAT CTTTAGTCCA TCTTAAGAA TTAAAAATGT CTTTGGGGA 300
ACATGTTTG TATACATAA TGTTTAGATA GAATATTTA TAGATATCTC TATGTGATTA 360
TTAATCTCCC TATGTATTT TATATCTAGA TGTGTCAATC TTTGTATTTA TATGAAATCC 420
TATGAATAGT GACA 434

(2) INFORMATION FOR SEQ ID NO: 31:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 715 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCAGGCTCC GATCTCAG CTCGACACT ATTCCAGCC ATACAGACC TGGTGTGAG 60
AAGCTTACT CCAATCTAG CCAAGATCC AAGTTTGT TCAATGGGG TTAGACAGCT 120
ATGACTATCT CCAAAATCA CCTCTGGAT TTTTCCGAG ACTTGTGTTT ATTGTGTTTG 180

CTGGCTTAT TGGACTCTT TTGGCTAGAG GTTCAAAAT AAGAGACTA GTGTATCCG 240
CTGTTTAT GGGATTAAT GCTGCTCT ATATCCACA AAGGCCATC GTGTTGCGC 300
AGTCACTGG GAGAGTTA TATGACTGG GTTTACGAG ATATATAGT ATAGAGATT 360
TGTGAGGA GAATTTCA AAGCAGAA ATGTGAGAA TTCACTGCA ACTATGAGA 420
AACTCCATG CTCTGCATC TTATCAATT ATAGTAAAC ATGGAACTC CATAGATTA 480
ATAGTATTT CTACAGAAA ATGGCAGAA AGTCAGTAT GAATGATTA ATTTGCTTT 540
CTTCTTAGG AAAAACTAG CCGAACTCT GTTATCTCT GTGAAATCAT CCTACAGCA 600
AACTAACTG GAATCCCTC ACTTAGAT ATGTACAGG CCTTAGACT CCTGATCTC 660
ATGTTCTAT TTATGACTT AATTAAACC CAGTTAAA AAAAAAAAAA 715

(2) INFORMATION FOR SEQ ID NO: 32:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GAGCAGTGC CGGAGAGG GAACTTCT CTACTCTCG CCACAGACC TGTCCGACA 60
CACTTCTGC CCTGCTCT GTGGAGGCC ACTTCTCCG CCAATGCTCG ATTCACCCC 120
CAGTCACCC TGAACATCG CCCCCTCT CTCTGCTTT GCCCCCTCT GCTCCCTGA 180
GACTGTCTG TCTTCCCTC TTGAAAGCA ATGCCAGCTT CCTGGATCT TCTGCAACT 240
CGACTACCA TGGCTTTGC TCTGTCAGC TCAGTCTCT AAGGGAATG TGTAACTCG 300
GTGTCCTCT TCCCTGCTC AACTCTCCA CCTCTCCA AGCTGGCATC TGCCTCCCA 360
CTGCAGAAA CGGTCCGCC ACCACTGCG TTACAGGGA GGAAGCAGCA ACATGAGAA 420
AAGAAATAT AAGGCTACA AAGATGCTA GCTCTGATC CGAAGGCAAA AAGATCTTT 480
GGGCAC 486

(2) INFORMATION FOR SEQ ID NO: 33:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 725 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GTTCCTCTCG TATATATAG GTTATTCOA GAGGCAAT GTCACTTTT AATTAAGAC 60
TTTCTGTTT AAGCTTTTC AAGAGCAG AGCACTTGA AGATTCGCC TAGGCTTGT 120
ATGTGCTAA TTCAATTTAT AAAATATAT CTGTCTTCA TTTTAAGCT TTGCTATAT 180
AGTCAGAAAT GTCTAAATA ACAAACTAT TTATATTAA TTATAGGAG ACTAAAGGGA 240
AQAIAAGGA AAATCAGTC TTATGTAG CTCAAGAT ATTAGGCTT AAGGCTCTT 300
TCTAGTTTAA TGAATTTG TACTACTGAT TTTATATAT TCTGTTTAT TATGATGAGA 360
TCTCTGGGA AATTTGTAG TTACATGCG ATTTCACTGT GATCCCTCTC AAGCTCAGT 420
CAATCTATA ACCAATGAC AACTGTCTC TTGCTTTAC TGTCTGTGA ATGTGAGCT 480
CAGTTTCCC AAGATCTCG TGTATATAT GATCAGAGT GCTTTTCTC GGTGGAGAG 540
TTGCTGCCC TCTTAATTT GTGTATGTG CTTCAGATA TCTAAACTC CAGTCTGATC 600
TGTATGCTT ATCTAACTG TTAATGTAT TATGATAT GTTCATATC TTGCTTAGG 660
GTTCATATT TTCATTTGA TAGAATAA GTTTTTTCT GCTTATAAA AAAAAAAAAA 720
AAAA 725

(2) INFORMATION FOR SEQ ID NO: 34:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CACACAGAT GCTGCCCTCA GACTGTCCA TCTGTACCA CATGAAGG CTGCTGCTC 60
TGCAGATAC TGAAGATTC AAGCATGCTC TCGAATGTT CCAGAGATC TGCAGATC 120
CTCTGCTTT TATGCTCT TGTGAAATC AATGTGAG ATCTTCACT CCAGCTGAC 180
CCAGCTGCA AAGATATCC AGTTCATCC CGAAGGACC AACACCGATC ACATGATC 240
AGGAATCTTA TAACTAGCT GAACTCTTC CATCTGACA TTGTGTGCA CATGCACTC 300
ATCACTGCG GTGCCAGAT CCTGGCAGG CAGACCTG TGTATATTC AGGTATCTT 360
CTACATCTG AGCTTAGCT TAGCTGATA TCAATTACA TTCTACTAT AATCAAAAA 420
AAAAAAAAA AACTCA 437

(2) INFORMATION FOR SEQ ID NO: 35:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 943 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

5 GGCAGAGCT GGCAGAGCA CTAAATCCCA GCAACTGC ATGTGTAAC ACATTAAC 60
AGAGTACTT ACCTCTGGA GATTATAT ATATGGTG ACATGATCA TGTACTGAT 120
GATGACCAAG ATGCTTAAG TCTACATTTT CTTTATCT GTTAATATTA CTTCTCTTA 180
TCTTGTTCA TTACATGCT ATTTCTCTT CAGTGTAT TTCTAGTGA CAGATCTTA 240
20 ACATTTCTTA CAGCTTGCA GAAGGAGAG AATGTGTT TGGGTGGGT ACTAAATTT 300
TTGAGTAAA TATCATAGA TGAATGGA AAAAGGAGA CACAAATGT TATTAACAAA 360
AACAATGCT TTTTATGCC ATTTAGTGG CTCTTAAAT AGTCTACAG ACATTCAGT 420
TTACATCAC TTTTATGGA ATTAATGTC CACTACTGT ATGCTCTTA AAGGCGAAA 480
TGTGCTTAG TGGCTAAGG CTAAATTTG GTCAATTTG ATCAGAGAT TTGTAGTAT 540
30 TGCATTAAT AGCACTAT TTTCATPAG TGTATTTT TGTATGAT TTTTATACC 600
ACTATTTGT TGTAGCTTT TTTGCTTAT AGTTGTAAT ATGACATGC TAAATTCAAA 660
CAGATTAACC ATTTGCGAA CTAAAGGAG CAAATTTAT ATGAAAGUA TTTTGTGAT 720
35 NOTCAATTC AACCTTGA TTAAAGCTTA GACTAATAG TAAATATAG TGGATAGAT 780
TTTGCTTTTG TGTATTTAT GTGATTAAG GATATGAT TTACCTTAT TTTGTAGUA 840
40 AGTGAATTT ATGTGTTTA TTATTAATTA AACTGTACC AGGAAAAAA AAAAAAAAN 900
NAAAAAAA AAAAAAAA AAAAAAAA AAA 943

(2) INFORMATION FOR SEQ ID NO: 36:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 604 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

5 GGCAGAGAA ATCTTCATC TGTATGACT CAGAGCATG GATGGCTTT CCACTGAT 60
60 GATTCCTATG TCTGCTGTC AGTGGTTGG TTTTCAATGT TCTTCTAAT TTTTCTTA 120

TTGATCTTG GAGTTTCTT TGTTTGCTT CTGTGTTTG CCACTTTTA TAAACACAG 180
CCCAACAAA AACCATAGA TTCTGAACA TAGGGGGCC ACATGGACC CAGTATGTA 240
5 CTTTATGGA CTTCAAGAA AATCTGAT CCGAAAAATG ACCTAGGAA TGTATACTCC 300
ACATTTTA TCCATATA TGGTGTGTT TCTTAATTT GTTTCTGTC GCGAATGTC 360
CCTTCAAT TAAATGACC TTTTCTCTT TCAACTTTT TTTTGTACT TGTATATTA 420
10 AGGTTTGA AGATTCATA ATTCAGAG AGTTTGA CAAGAGATC 480
TCAATGATA TCTTTTCAT GTGCTTTAC ACCCACTAC ATTATAGAT GTATATGTA 540
15 CAGAAATAT ATGCTGTGT ATGTGTCTT ACTCAATAA GTACATGCTT CCGAAAAA 600
AAAA 604

(2) INFORMATION FOR SEQ ID NO: 37:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 349 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGATGCTC GGGAGCCCG AGGCGCTGC CTAAGAGG ATATCTTTA CCGCTGCTT 60
GTCCAGACC TAAACCCCA GCTCTCAGG CAGTGGCAC ATGCGAGGG CTTCACTGG 120
GGCACTAGA GCAATTTGG GACTGCGAT GCTCACTTT GACTTCTGC AGTTCGGGG 180
AATCCAGAT CAGTATGAC AAGTTTACA TATCTGAT CTTCAATGT CTAATCTGC 240
40 CCGCTCGGG TCTACAGG TATATGCC CACTTCTGT TCTAAATTA GATTAAGAT 300
CAGTACAGA GAGCACTGC CCGAGAAA TCGAAGCG TTGCGACA 349

(2) INFORMATION FOR SEQ ID NO: 38:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 672 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GTATGCTTG CCGTTGGCG GATGCGAG ATCTGCGCT TTGAGCTGT AAAAGCCAC 60
TGGTACCGG TCGTTGTTG TTTGTGAT GTATGCTTG CTACAGAGT GATGCTGCA 120

5 GAGAGCGAG CAGCGGTAC AGCGAGTAA ATAGTAATGC CGAGCGAGT TTCTCTCGGC 180
TTTATCATGT CAGCAGCTGT GGTATATGCG TTGTGTCTCG CGAACTTTGC CGTGAACAT 240
TTTACGATTA ATCAGATGCG GCGTGGCGCA ATATTCAGTA TACGCGCTCG CAGTGTGCG 300
GCTGATGTGT CAGTGGCTGC GCGACGCTTT TTGCGGTATG TTGCGACGA GGTCTTTTAA 360
ACAGTTTTCG SAGCGGTATT AGCTGACGCG GTTCAATGCG GGTGCGTAGC TGTGCTTAG 420
GTTCACGCG AGCATPAGCA TTAAACATCT CATCAATTTG GTTCTGCGTG GCGGTATCA 480
TACTTTTCAG CATATTTTAA GCGTGGCGA AACGCGTTAG CGTTTGGCCC ACGGTTCAT 540
AGCGATGCG GTTATATGAG TATCAATATA CACGACAGC TACGCGCTTA GACACGTTT 600
CGATTCGCT GCGTGCAGTG GTTAACACCA CGTGGCGGCG ATATGCGCG TGCACAGTT 660
CATGAGTAA AT 672

(2) INFORMATION FOR SEQ ID NO: 39:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

5 AGATGTGTA TTTTATGAA CAGTATTTT ACTTTTAGG AATTTGCTA GCTCTTGAC 60
TTNAGAGCT TTAGGAGCTC AACATTTCTT TGTAGAGAC GTTCTTTT TTGATTTTA 120
CAGTATPAA AACATTTGTT TTGTGAAT GTATAGTGT AAAAAAGGA TACTGTATG 180
CAGGTTTCAA AAGGAATGT GCTTTAGGA TGAATCAAA GATGCGATG TACTGTAGG 240
CATTTTNTT TCTTTTNGA ATGACATCA GCTCTTCTCT TCTGACTGT AACACATAG 300
CCCAAGCAT GAGTTATNTT TTCTATGCGT TTTTATTTT GTTTAGTTT GTTTGTTAC 360
GCGAGCGAG TGTCTCTGCG GACACTGAC TCTGCTCTCT ATGAGACA AAGTTAGAA 420
TCTCGGATA ACCTAAATA ATTAGAAAT GATTAATAA TGTGAATCG GGTTAAGTG 480
ATGATGATA ATTAGCTGC AAGAACAG CTTCTTCAT CAGACTGCG TACTGTTTC 540
TTCTGTAGC ATTTGTTTG GAGAGCTC TTGTTTCTT CTCTTTGCGG TATGCTGCG 600
TTTCTTATA TGTTTGTAC ATTTATGAG TATTAATCAC ATACCTTCA ATTCATTTT 660
TTTAGGATA CAATTTAGT GTTTTATG TATTCAGAA GTTGTATAC CGTAGGACA 720
GTCAATTTTA GACATTTG TTACCCAAA AAGAAAGCT GTACCTTCA CGATCACT 780
CTCATTTCT CCGAGTGGC ACCGATGCC CGAGCGCGCG GAACACTAA TCTATTTCT 840

5 TCTCTTACA TTCTTTATT CTGTCATTT CATTAATATG GATTTCTACA ATATTCGCTC 900
TTTTCGACT GCGTTCCAA ATATGATTTT CTATATGAG TCGAAATTT CTTCATCT 960
TGAAGACTT TATTCGTGTG AAGCGAGTG GTTGATAAA TCATAGATTT TCGCGCAGA 1020
GCGCGATA CTTACAGGCT TTTTCTCGT GATCTTATG CTGAGTAGTT TTTCCTATA 1080
ACCAAGCAT TATCAATAT TACTACTTAT ATATCTATG CTAGTCTCTA GATTCATGT 1140
TGAACTTTT GCGTCTCGAG TCGGAGAGG TCTCTCTAAA AATCAGCTA AAAATCAGC 1200
CAAAATCAG GCGAATGAC TTGCGAATA ATTCAGAAAG TCGTTTTCAC GTTGTCTAT 1260
CTTTCTAGC AGCTTTTATA CTTAGGCGCA GTTAGGCTCC CCAATTTTCT TTTTCAAT 1320
ACTCAGTGA GTTTCTGCTG TCTTTTCAA GTATATGCA TAGGACTTAA AGTGTATTT 1380
GATGCTTGT AACACTGCTA ATATCTTAA GTACAGATTT TATCTACAG TACTGTGCA 1440
CAGTCAATTA TTGCTTAGGG TAGTCAAAA ATATGATGAG AGCTAGTTAA GCGTTTCTT 1500
GACTGATTC AGTATATTC AGAATGTGT ACCATCAAG GCTCTTTTAA ATACGAGCG 1560
ACTCACTTA TACCGAGGA ACCAGCGAA TACTGTGAG CCGCAGATA TCGATATCA 1620
TGAATTCAG GTATATATC TCTGTACTC CTAATGATTT GTTTTCTAAG CATGTGCT 1680
TCTCACTGCG TTGACAGCAT CTCTGCTGT GTATGCGCC TGTTHACATG ATGATTTGA 1740
TATGTGTTT TGTGTGAGC ATCAATGCT GTAAACCAA ACTAAACAG TTTTCTTGG 1800
ATATGTTCC ATCTTTTAAA TGACCTGCC CTGTCAATA ATTAATGAT TGTCTACCC 1860
TGTAAATAA AAAAAAATTT AAAAAACTG GAGCGCGCG CCGTACN 1908

(2) INFORMATION FOR SEQ ID NO: 40:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

5 CTTCAAAA AAAAAAGAA GAAAGAGT CTTACAGAA GCGGTGATT CTTCTGCGA 60
AGGATPACA TCGAAATGT TCAATTTTC GCTATATGT TCCATGCTT TCCCATCCA 120
GCGTAAAGA GAAACAAAG ACATGATCG TATCTGTGT GTCTCAGCT TTGCGACTTT 180
TGTGATGTT CTAAGAGAG AGTACCTTG CTAAAGAGC TCAATAATCC ACCCACTAA 240
TAGCTAAGT GCGGAGAAA TGAATTTTC CTTGTGAT CTCCCAAT CCAATTTGT 300

CACACAGCCC TCCAGAGCC TCCCTAGTTC CTTCACAGTC CAACTCTGTC TACTTGGCCC 360
GCACCCGAT ACTATGTGTC CTACATTCAC CTTCACATGG CAACTGGCTT CCACTTCGCA 420
CATMAAACCT CATATTTTAA ATTAAGTTGA AATTTGAA 458

5

(2) INFORMATION FOR SEQ ID NO: 41:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1153 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GCACAGAGC CTCGACCCA GGTGTCTGTC AGCTGGCCGC GAGATGTGTC GCATCTGAGA 60
GGCTGTCTCT GCACTGTGCT TGGGGGAGGT GGGAGCTGTT TTAACTGTTT GCTCCCTCTC 120
CTGTGCGGCC GTGGCGATCC CCGGGGCGAG TGGACCCGCG GCGTCTCTCC AGCTTCCGAG 180
TCCAGCCAGC CTGGGGGCGG GCGCGGCGCC GAGACACCGG AGAGTGTGCT TCTTCCCTCG 240
TTAGCTGGAC TGTGTAGCTG GTCTCTGTGT GCTCAGAGCC GTGGCGAGGT TGAAGCTTAC 300
CTGGCGAGGT GCGACGAGGG GTTGTAGAGG AGGAGGAGG GCHTGGAGG AGCTTGGAGA 360
ATCGGTGCTC CAAACTCTAC ACTCAAGAT GCACTGGCCA ACTCTGTGTC GATGTGGCTG 420
GGGCAATGTT CTTTGAGATT CTACACAGAG AGAGTGTCT GCTGGCCATT CTGACGACGAG 480
TGCATGCCAT GCGAGTCTG GAGGTGTGAG GTGCATGTG TACGCTCTGC CAGCGAGGCG 540
GAGCGGCAGA TCTGCCGCGA GAGGTGTGCT GAGAACTCT GCGAGAGAT CATGACATTC 600
GTGGAGGTGA TGAATCGCA TGAATGACTT CCGAGATGC CACACAGTTC GAGGTGTGAT 660
AACTGTGTTG AGACAGGCTT CCGGAGCTG CAGCGCTACC TGTACAGAT CTCTTTGCGAG 720
ATCACTGATG CCGTGGGCAC CTCACTACAC ACCACATGC GCAAGCTCAT CAAAGACACC 780
CTGCGCTCTG AGCTGCTCTG GATCTCTGCG AGCTCTCTGA TGGCTTCCAG AGCTTGGCTT 840
TTGGGAAATG CACTTTTGGC CTTTGGGCT CTGGACCTG CTCTGGGTCA TTGTGTGAGAC 900
TTGGAGAGGG CAGCGCCCGC TGGTCTCTG GTTTTGTGCT TGGCAGCTTC AGGTGATCTT 960
TTTAATCTTT GCTCAAGGTT CAGTCTGCG TCTACTGTCT CCGATAGCC CTGTGCGGCT 1020
CCCTCTCTCT TCTCACTGTT AGAGAGAGC CACACTGCG ATGGCGATA AAGTTGAGAA 1080
CATGAGTTTG GCGTGAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1140
AAAAAAAAA AAA 1153

60

(2) INFORMATION FOR SEQ ID NO: 42:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1983 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GCACAGAGG GGGCGAGCC GACAGATGT TCTTCTGCC TCTTCCGCTT GCGGGGCCAG 60
TATGTCTGCG AGCTCTGCC CTGAGAGTTT TCGGAGCGCG GATCTCTCTC ACCCGACACA 120
TCAAGAGCG CTTTGTTTTA GGAATCTATT CCAAGAAAA AGAGATGAT GTGCCAGAT 180
TCAAGATGC AGGAGAGAT TTGATTAAT TTATTAGCTG AAGCTGAGA GAGACTTTCA 240
ACTATCTGCG ACCACTCTG AAGGCAAGCA AGACTCGAC CTTTATGCTT CTGCATCAGG 300
ACTTCCGAG GGTGTGCTTA GTTGGCTCTG GCAAAAAGGC AGCTGGAAT GACGACAGG 360
AAAACTGCA TGAAGCGAAA GAAACATCA GAGCTGCTGT TCGAGCGCGG TCGACGAGA 420
TTGAGACTT GAGCTCTCTG TCTGTGAGAG TGAATGCTTG TGGAGAGCT CAGGCTCTG 480
CGAGGAGGC GTTCTCTGCT CTCTATGAT AGATGAGCT AAGCGAAAA AGAGATGCG 540
CTGTGTGCG AAGCTCTAT GGAATGCGG ATCAGAGAGC CTGGGAGAAA GGAATCTCTT 600
TTGCTCTGCG CGAGAACTTG CGAGCCCAAT TGAATGAGAC GCGAGCCAT GAGATGAGC 660
CAACCGAAT TCCGGAATT ATTGAGAGA ATCTCAAAAG TGTATGATGT AAAGCGAGG 720
TCAATATCAG ACCCAAGTCT TGAATGAGG AACAGCCAT GGAATCATTC CTGATGTGCG 780
CGAAGGATTC TGAAGAGCCC CGAGTCTCTT TGAATATCA CTACAAAGGC AGCCCAATG 840
CAAAAGAAC ACCCTGCTG TTTGTGCGA AAGGAATTAC CTTTGAAGAT GGTGTGATCT 900
CGATCAGGC TTCTGCAAT ATTGAGCTCA TGAAGGCTGA CATGGAGAA GCTGCAACTA 960
TATGCTGAG CATCTGTCTT GCTGCAGGC TTAAATTTGCC CATTAATATT ATAGCTCTG 1020
CCCTCTCTTG TGAATATGTC CCGAGCGCA AGCGCAACCA GCGCGGGGAT GTTGTTAGAG 1080
CGAAGAGCG GAGAGCAATC CAGTGTGATA ACATGATGC TGAAGCGAGG CTGATAGCTG 1140
CTGATGCTT CTGTTAGCA CACAGTTTAA ACCCGAGAT CATCTCATAT GCGCGGCACT 1200
TAAAGGCTG CATGATGTA GCTTTGGGAT CAGGTGCGAC TGGGTCTTTT ACCAATTCAT 1260
CTGTGCTCTG GACCAACTC TTGAGGCGCA GATTTGAAAC AGGGAGCTT GTCTGAGAGA 1320
TGGCTCTCTT GGAACATTA ACAGACAGG TTTAGATATG CCACTTCTCT GATTTTACA 1380
ACATTTGAAA ATAGATCTT CCGAGAGCAT GTACAGCTGC AGCATCTCTG AAGGAATTCG 1440

60

TAAGTCAATCC TAAGTGGCA GATTAGACA TAGAGGCGT GATGACGAC AAGATGAG 1500
TTCCCTATCT ACCGAAAGGC ATGACTGGGA GGGCCAGAG GACTCTCATT GAGTCTTAC 1550
TTCCCTTTCAG TCGAGCANT GCTTAGTTCA GATACTGAA AATGCTGTCA CTCGTGCTTA 1620
AATTGGCAG TTGACTTAA AGGTTTTS ATTAATGGA TGAATATCTT TTACGGAGA 1680
CAAGAGATG TAATTAAAA TGTAGACAC ATGAAATTT GTATGCGCTG ATTTTITTT 1740
GATTTCAC AAGATATAT AAGGTAAAG TTATATCTT ACTTGATAG GATTITTAAG 1800
ATACTCTATA AATGATAAA ATTTTAAAG CTTCCTATAC ACTTTTCAG GTATATGTTT 1860
TTCAATGAGA AGCAAAATG TAACTAGAT TTGTATGCT AGGAACATCA GCAAACTGA 1920
AATTACTATG CACTGTGTCG AATGATAAA TGCACCTGCT TGTGCAAAA AAAAAAATA 1980
AAA 1983

20

(2) INFORMATION FOR SEQ ID NO: 43:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGATGATCA CTTTGAGAC GATTATATC CTCCTCTCC ACCTAAGGC CTTGAGTTA 60
ATAGTTGGA AAGACTATAT AAGATATGAC ATTTCTTCA GCGAGAGGA AGATCATGT 120
TTAAGGCTTA ATGCTTACG NCTGTGCTT AACTGACCT GCGATTACTA CTTTAAGAA 180
ATAAAAAAT CAGTCAAT ATTCCTGAC TGAAGTTTA GTGGCAGCA TTCTATATGC 240
CCTTCACTTA TCAAGCTACT ATGTAGAAA GTGTACAGCA TACTGACTCA ATTCTTAAGT 300
CTGATTTGTG CAATATTTA TGTACTTTT TAAATAGCTT TCTTAGCTGC AATTCATGAT 360
TAGAGTAAA GCGCTGTGCT AATATAGG CTAAGCATA ATTGTACAT GATAGCACT 420
TTCCAGAG GAGCTTGAA ACGTAAATGT GCGTACAGG TTATTTTAC TGTAGAGCA 480
TCAGCTGCTT CTTTAATTA TGAATTAACA ATATTTTAA ACAATATATA GTAGACAT 540
AATTAGGTTT TCTAGTATG TAAATGACC ACGATGATC TTTCGCTTTT TGAATTAAT 600
GCTAGTGTG TCCCCCTGG AATTGTGCA GTTTCACAT GTTTCCTGCG CAGGTGTAC 660
TGTTTGTGCG CTTTGTAA ATGCAGACC ATTTGTTGG AGTCAGATG GTTTCATTA 720
AAAAAATAA AAAAAATAT ACGTACAGC TCACCTTTCA GTTCATATA TGTACCGAG 780
GTAGCAATGT GTGGATGAG GTTCATACA GACCTATTA TTGCTTGTCA TGTAAATTA 840

60

AAAGCTTTGA TTAACTCTT TTAAATCTT TTAGTAAAA TTCTCTTTG CAGAAATGAT 900
TGTGCTTAT TTTTCAAAA ATTGCTGTG AACAGCTGA TGCAGCAG CAGATTTAT 960
CTATGACAT ACAGTATCT TAAATGCTT CTTAGTATT TCTGTCGAC TTGTAAATG 1020
CTACAGGA TATTAAAAA ATCTATCTAC TTAACTTAT ATAGTTTAT GAATTAATAA 1080
CATGATGAC AGCTTTTGT CTGTGTAAC CTATAAAAA AGTTTGTCT TGAATTTCA 1140
TGTAAAGAC TGAAGCAAT GTATATGTT TAAATATTT TGTGTTGCA GAATTTTTC 1200
TCATAGAAA TTAAAGAAC TTACAGGAA GGTTTTAAAG TTAAATAT TCAATGCCA 1260
TAAATAGA AATTATTAAT ATATATTTT AAGCTGAT CAGTGGAGT CTTGCTATG 1320
TAGTATGTA GTTATATAT AACACAGAG ATTTTITGG CTATTTACG TAACCAAGG 1380
GGCGATTA NTGTTTGA GACTTG 1406

(2) INFORMATION FOR SEQ ID NO: 44:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1391 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGCTTAGG GCGGCGCC AGTCCGAGC AGTCTGCTT CTTCTGCGG GCGCTCGCG 60
CCCGGCGTC GCAATGACA GTGAGCTGA CATCTTCTG GCGACAGCA CCGTTATGA 120
CGAGAGTG TATCGCTCT GCTCGATGG TTACTGCTG ACAGACCGG TCGCGCTCG 180
GCTGCGCTG GGAATCTCG ACAGACTCG CCGACGGCA GCGTCTCTG AGAGCGAC 240
CATGACAT TACCGACCT TCCATGCT CAGCGCGCTG CTGATGCG GCGCCAGCT 300
ACTGCAACG CTAATCTTC AATTTGCGC CTCGCGGAG GCACTACTA TCGAGAGTTA 360
CTATGCTTT GATAGGCTT TTGTTGCGA GGTCTGCGC AAGAGCTGT CCAAGGCGC 420
CAGGAAGAC CTGATGACA TCGAGCACA AAGAGGATC ACCCTGAGA GCTGCGCGAG 480
ACAGTTTGA AACTTAAG GGTCTTCA GGTGTAGAG GAATCGCG GCTGCTGCT 540
GGCAATAT CAGCAACT TCTCTCTCT TGCAGGCTG GCGAGGACT ATCAGACAT 600
GCTCTCTTT GCTAACAC GCTTTGAG AGGGAGAAA AACTGCAAT ATCTGAGCTT 660
CGTGAATTT GCTTCTCG CTAAGCTAT GATCAAAA TGAACCTTG CACCGCTGA 720
CTCAGAGTG GATGACATG ACATGCACTT AAGAGCAAT TTCTCAGCA CTTGAAGAG 780

60

CTCAGCTGTC TAGTGGCTCA CAGAGACCTT CTGACCTGTC ACAGAGCCTT GGTGTGACT 840
GCTCTCGGG AAGCTGAGGC GTCTTCTG AGATGGAGGC CAGCTTGAG AACCTGTGTC 900
GGGGGCTGCT GAACTGTGCG CAGAGCTGAC CCAAGATGAA GATGTGAGAG ACCGTGTTGT 960
GACCTCTGTC GAGAGTTTTC TGGAGCTGTC CCGCTCGGAC CAGTGGGAC TCAGAGAGCT 1020
GGGTGTTTTC CTGATCTACT ATTGAGCTTC TGTGATGTC CTGATGCTT TCAGAGACCA 1080
GGCTCTCTG GACGCTTACA TGGGACCTT CCGGGCTGTC GTCTGTGCG TGTATCTGTA 1140
CTGAGTGGC TCCGACCTTC GCGGACCTGTT GACATATGAG TTGCTCTGAG TTGTGAGACT 1200
GGTCTCTGCT CCGGGGAGCA AGTGGGGGTC GTGAGATGTT GCTGTGTGCT GTCTCTGAGC 1260
ACCTGTGTC CTCTGACAG GATGATGTC TGTGATGCT CTTTGGGAG TCAGAGATAT 1320
CTCAGGAAAT GGTGTCTGTC CTGAGCCCAT CACCGAGAG AGTGTGCTCA CAAAAAATA 1380
AAAAAAAA A 1391

(2) INFORMATION FOR SEQ ID NO: 45:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1569 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GGCAGAGTC GAGATGCTG CCGGCTGTC GGGATGTC GAGGGGCTC TCTTGGCCCA 60
GGCGGGCGCG CTGCTTACC TCAGACTGT CCGCTATGTC TCAGAGCTG TTACCGGCCA 120
CGTCTGTG ATGCACTTC AGCGGAGAA GCTGATGCT GTGACTGAT ATATCGCCCC 180
GAAACAGCC ATCAGCCAT CATGCTGTC ATCTCTGTC AGCGGCGGAC AGGAGGAGAT 240
AGGCTCATC AGCTTCTG CCGGGAGAT AGCAGAGTT TTGAGGACA ACCGATGAT 300
AGCGTCTG CAGATGTC CTGATGTC AGAGAGCAG CTCTTATTC GAGACGACG 360
TCCGGAACA CAGATGCT ATGAGTGT TCCGACCA GTCCTGAAA CCGCTCTGTC 420
GAGGATCCA AGTACAAA TCTGCTGTC CTTTTGTG GGCAGCAT GCTGCTGTC 480
AGTAGAGGC CAGAGTCA GAGATGTA CGATCTTAA GGAATGTC CATCTGTC 540
GCTGTAGAT GGTGATGTC ATGAGCAT CTTCAGAG CAGGGCTTTA TCACTACTC 600
CAGCTGCCC AGCTGCCC TGTGTCGCG GAGCTTGTTA GAGGGCTTCA CTTGCTGTC 660
AGCGAGACC CACTGCTG TCAGACCA GCGGCTGAG CTGACACC TGTGAGCA 720
GTACATCAGA GAGAGGCG AGGAGATT CTGTATGTC GCGGATGTC AGCGGATC 780

CTCAGACTG TCCGACTG TACGAGCT GTTTAGCAG CCGTGGCAT AATATCACTC 840
TCCGTTATG GCTGTGCTT CCGATGCG AGATGTGAA GAACTTGGG TCGGGGAGTG 900
TGTGTGAC TTGTTTTC CTATATGA TATTTGAG TATAGGCCA CTTCAGATG 960
CAGAGATTC CATTTGAT GTCATGAC GGTGTGCTC TTATTTTTC CAACTTGGGA 1020
CGTATAGA GCAGATGTC TCGATCTTC AGGTGAGG CAGAGTCTT GAAAGATGAG 1080
GCTATGTC CCGTGGCTC TTGCTGACT GCTGTGCTC CAGGGCTTC TCAGGCAAC 1140
CGCTTGGGC ACAGCTGTC ACTGAGAG TACGTGAC AGCAGTCTT GCTGTGATG 1200
GAGCTGTC AGAGCTCT GTTGCCAG CTTTGTGTC ATGCTGTTA CATGTATGAG 1260
TTGCTTACT GCTGACCA TTGTACCAT TTGAGAGG AGGAGCAG AGATTAAGTG 1320
GTTGTGCA GGTATGAG TTATGATG GAGAGAGG GACTTGAAC AAGCGCTGTC 1380
CTCTGAGAC CCGCTCTGA ATTCTCTAC TAGGCTTTC TCATGATTT ACCGAGAT 1440
GGTTCGAT CAGCATGAG GTTGCTGTC ATGTTATTC TCAGGCTTC AGGTGATGTC 1500
TGTGAAAG TTGGAGGAC TCGTTATTA TAAATGAAA TATATCTTA AAAAAAAAA 1560
AAAAAAAA 1569

(2) INFORMATION FOR SEQ ID NO: 46:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1924 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GGGGCCCCC HCGKTTTTT TTTTTTTTT TTTATTTAG ATATGCTTT TATTATCAG 60
AATGAAAGT TCATTCCTCC TTCACTCTT TCTCTTTGT TTCTGGACA CAGCTGACT 120
GATCTGCTA GAAAGTTGT CAGTCTGCT GTGCTTCCC TCTGTATG ACTCAGCTG 180
TGTGATGCT TGAGAGATAT CTATCACTT CATGTGATG AGCAGTCCA TATCAGCCA 240
CATCATCAT TCTTACCTTA AGATATTA GAAATGTTA ATATTAAGA CAGGGTTATA 300
AATTAAGGT TTGAATGTC TATCACTT CAAATTTTAA AGATTAATA TCCAGATTA 360
AGATTTGGG GTATGTTACA GATTAATAA ATAGAGAGA ACTTATGTT GGGGGGAAA 420
TCTAATTTA TTCTTACATA AATTAATAG ACACCTGAT TAGAATGAAA ACTGTATTTT 480
CTTTAATG TAAAGCTG ACTCTGAT TTCAAGTCTT GAGGAGAT TTGACTGCA 540

5 600
660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
1380
1440
1500
1560
1620
1680
1740
1800
1860
1920
1924

CCGAAATAT ACTATCCCTT ATGTGAAGGT ATGTGACAC GTTGAAGCTA CCAATGAGT
TTTAACTCA GCTCTTTTTC CATATGAGG CAGATACCTT GCTCCCAATT CAGATATGTC
TTCCATGTC AGGAGGCTG ACGACTTCA GAGAGATCC TCGAGAGTCC CCAACTGCC
CTTCCACAG TACACACGC TGTAGTCTTT GTCTGCAAT CTTTGTATTT TACTCTATTC
TTTCCATCT AAGTCTCAC TGAATTTTAA AGTTAGGCTT GGAAGACTA TCCCTTACTG
GAGACGAG GAAACATTT TTTTCTGAGG GAGAGAGAT TCACTTTCAC TATATGCTG
GAGGGCCAC AGTGCACAA ACAAGATCA GCTTTCATTC AAGTTCCAGG TTTTCTCTCC
TCCCTGAATG ATTAATGCA AGGATATAT AATTAAGAT TCCCTGTGTC ACATGTACCA
TCAATAGG ATATATATC GTTTTGCAAT CTTCGCCCA TTCTCCAGAT TGTCTATCT
TAATCCAG CCGTTTTCAC TGTCAAAA AAAAAAAA TATTTTTTC AGCACTGGTG
TTCAAGCA AGTTTTTAT GTTATAGT TTACAGCA CTGTGAGAT TTCCAGTTGA
GTCTAAAA TTGCAATCA TTATCTAGA GCAATGAC ATGATTTAGA GCAATCAAT
CCTCTGAT CTTCCTTAA TAGGACCA TTGTAGACT GCACTAGCTG ACATCACTAA
AACATATCA CTAATGCCA AACCAATA AAGGCCAG CCAACATCT GCTCTCTTAA
AAGCTGCA AATATCTTAA GTACAGCA GTAAATGAG GACTGTGAGT CATGTCACTG
CAGCTGACA TGAATTACA TAGAGAGAT GCAACGCCA TTAAAGTTAA AAGTCCAAA
CTAGTCAC GATCTCTTT ATTTGGGAAA AGTGAGACTA TTATGCAATC TTGTAGGTT
TGCACCTG CATGAAGGC ACCAATGCA TTTCTTTTAT CTTTGAGAA GCAAGGTAT
CTGTTCAG GGCCTACAG TAGGAATTA CATCTTGGCA TCACTCTCT GAACCCAGA
CTGTCTCAT TAAATATAT TTTGTCTT ACACAAATTA TGAATTACA TGCAGCAC
TGGATAGC ATATTAATC AAGCACTTA ATTCACACT TTTTGAAGTT TTTTAAAAA
CCCACTGAC TAGATTCAC AATCATTCG TACATACAA TTAAAGCTAG TAGAGACCA
CTAAGCTG AGTTTCTA TTTTAAAGT CAAAGCCTA ATCATCTGAA AGTGAACAG
GTAA

(2) INFORMATION FOR SEQ ID NO: 47:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 475 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

5 60
120
180
240
300
360
420
475

TGCTGTGGG CCGAAMAC AGGACACAG TGAAMACAC CCGAGACT TGTATCCGC
AGGAAGCCA TTCCATATTC TGAGCCCTTG AGGCCACAG AGGAAMACG TTTTACCGA
GCCCATGAG AACTCTCTTC ATGAGAGAG GGCACCTTG TAGAGACAT CATTAATAC
AGACTGTGG TCCCAATG CTGCTGTCTC TCTTTGAGA ACCACACAG ATACATCTCC
TGAAGACCA GTTGGGACA GAGAGGCTG GAGAGAGAT CTGAGAGAG AAATGAGAA
TAAGCAGAC AGTGTATTC ATTCTGTAAA ATTCTATAT AGAGGCTCA GTTGTAGAA
TAAGTTATTT CTACTAGTGT CAGTTAAGT GTTCTGTCTT GTTCTGTCTT CCTGTTAGA
TAAGTAACT CCGTTTGGAA CTACAGAGT ATGTCTCTCC TTCAATATGT GTGAA

(2) INFORMATION FOR SEQ ID NO: 48:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

5 60
120
180
240
300
346

AGGAGACAG GACTGAGT CAGATCTCAT TTACAAACA AGACCCCAT GCGAAGATC
ATGTCTGAAA TTCTAGCTT ACTGTCTCC CTGCTGGAC CTGCTCTGGA TAGAGAGG
GAGGAAGAG ACTAATCAGA GGAACCAATG AAGTCACTCC ATGATTTTCC TGAACCTCC
CGAGTAGAG ATTACGTTT GACCTTCAC GTAGAGACT GTGAGATG CTACTTCTG
GCGCAATGA AGACCAAGC CAGAGCAAG CCGCAAGCTT GCTAACAGG CAGATCTTT
GCGCAGGCA CTCTGTGAA ATCTGCTTC CCGCAGAGC TGAACC

(2) INFORMATION FOR SEQ ID NO: 49:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1366 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

5 60
120
180

TAGTGTCTAG CCGCCACCC CCCCCCAT CAGATTTAC TGGCATGCT AGTGCCAGC
TTCTAACAA GCTGTATTT CAGTCTCTCT GAGACTCAC TCAGGATCA TACCCCTCA
GTAGAGCAG CAGGTGATCT TAACTCTTT CAAAGAGAG CCGTCTCTG GAAGCATTT

5 CTTGACGAC CACAGCAACC CCTCTGGAAA TGGATGACAA ACTGACTTCT CAGCGAGCA 240
GCCAGCTT CTATGTGTAC AGTAGGACAA GTATAGTGG ATCATACAT CAGCTGGCTT 300
TTTGGTTTAG TCATCTAGAG TGTCTGGAC TAAAGTCTTT TCAGGTCTCC TTGCCCTGG 360
ACTGGGTGAA CCTGCCGACC CGAATTTGCT CAATTGTCTCT GAGGCTCATG TCTCTCTGG 420
TGGTGGGCCA GGGCCCTGCA TGGGAGGGA CCTCTCTGG GGGGAGGCA GCTGGGGTG 480
CTCACTTAG CGCAATGAAA GTATTGTAG GACTGTGTGT TGAATTTGCT GAGGCTATCC 540
TTGATGGCA GGGGAGTGC GGGAGTGA CCGAGTGTCT GCGAGGCTCT TCTCTCTGAC 600
TTTCTTTCT GTCCAGGGG AGCTGGGAGA AGCTGTGAG GGGGCTGAC TCCAGCGCC 660
TCTCTCTCT CCTTCCCTGG CAATGTCTTG GGGTGGATTT CCTTACAGCT AGATTTGCA 720
GGCTTACTT TTCTCTCCAC AAGAGATGCG CAGCCAGCT AGCTCTGACT TCCCACTGTG 780
ACAAATTTCA GTTAGAGCT CTAGCGAAG ACTGGGCAAT TACGCGAGG AGAGGAGCT 840
GTGAGTCTGA CCGVAGSGG GRCGCTTCA CCTTGGCTGG CCTGTCTCTG GTCTTAGGT 900
TTTCTCAGT TGTCTTTT TGTATCTT TTAAGCTAT CCAATATATA GGGCGTGGG ATGTTGTAG 960
CGGCGCTGG AGTGTTTCT TTAAGCTAT CCAATATATA GGGCGTGGG ATGTTGTAG 1020
AGATATGCA GATATGCT GTTTAGAC CAGAGCTTGG GACCAAGGCT CCTACAGCTA 1080
ATTTCCTCT CTGTGAGCT AACAAAGCT TAAATTAGCT TAACTAAGA ACAGGCTCC 1140
GTGACGAGA GTTCTGAGG CAATCTTTC AGTTTCCCTT GTTGCAATT GCTCTCGAT 1200
TCTATGAAA CGACAGGEC TTAGGGGGCC TGGCCAGGA ACGCAAGCT CTTAGGGCTG 1260
AGCTGTGAC AGCAGGGGCT TGTGTCTCTG TTCTGTCTCT CTGCTTGGCG AGCTTTCTCA 1320
ATAAAGCTTA TTCTTTATTT ATAAAAAAA AAAAAAAA AAAAA 1366

45 (2) INFORMATION FOR SEQ ID NO: 50:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

55 GCGATATTC CTGTAGCA CTGATCCAC CAATACTAGT TTATTTTTC CCTCAATTC 60
ATGATTTTAA CTTCTGTTAC AAGGGGAAT TTGCTGATAG CTTCTTGGCT CCGACTGTTC 120
CAATTATGCG TAATGATTC CATTTAGGG CCGAGGCTCT TCTTGACTGA TGGTGTGTC 180

5 TTTAACCCCT GGCATGTATA ATAGAAATTT GGTGAATGAA AGAACCCAAA TAGCCCAAT 240
AGTCCGCCCA GGCCTGATA TCCATTAAGG GCTTGGGAAT GCTATATGTA ATTCTCTTTA 300
GTCTTTTGT TGTTTTAAA AAAAAAACA AGATGGCTC AGATGGATGC CTACGTAAAA 360
ATGGTTCTTA CTTCTGACT CATTAATTTT CTTTGAATTT AGTATGAAA GGAAGAGGA 420
GGAAGGAAA TTAAATGTC TCTAGTAT CTCTGACTC AGTCTGACA TATGAGTAA 480
TAACTATAT TGAATGCA AGATTTAT CTGAAGAG AGTACATTT GACACATTTA 540
TCATGCTTC ATATTACATA TTAAGTAAA CCAATATATA AACTATGCA ATATCCATTC 600
CACAGGCA AGGAGCTTA ACCTTTTGT TCTTTTCTA CATAGGAA ATGATTTT 660
TTTTTATTT TTAGGGGAA CTTATATAT TATGAGCCAG TGAATCTTT TGTGACTTA 720
AGCTTATGAA TTCAAGTTAC AATGAGTTG ATCTAGATG GTTACTACT TGAAGAGAT 780
GTTGTGCTT TATGTGAC CAGCGAGGC CTCTGGGGA ATTAAGAGG CAGTTTTCAT 840
GCCAGACA ACTGTAGCT TTAGTGGCA GATGGGAGT GATTACAGA CTGCCAAAA 900
TGTGGGGCT TTGGGATTT CCAAGCATC CAGCTGTCT TGTCTATCT TCTCTTTTC 960
ACACTCTGG ATGCAATTT TGAATGCT GRAMATTCY TTGAATTTT CCCAATAGCC 1020
WTRGCCAC ATCTTATG ACACATAC CAATAGATC CATTATGTTG AATATTTAGA 1080
AACCTGTAC CAGCTTGA KTAACCGMA WTTTTCCTAT GTTTGTGAA TTGATTTGA 1140
AATACAGGG CTAGGAAT ACTGGAGT TTAGGCTGT GGTATATAC TTAGGTTAT 1200
TTAAATTTT GTATTTTAT TTAAATGTC ATGATTTT GAAAGAGCA AATTTATCAG 1260
GATGGCTCT TTGCCATGG TCTTATTTT ACCTCTTTT CTTTANGAA AAGAGCAT 1320
GTCTTATGT ATTTTAAG TTTTGTGAT AGTTTCTAT TCCATTTTA ATAAAGTTT 1380
TATTAATAA AAAAAAAA AAAA 1405

45

(2) INFORMATION FOR SEQ ID NO: 51:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

55 GCGATTTCT AGGACCCCA AAAAAAAA AGGGAAAA AAACCCCAA ACCGCCAA 60
AACCCCAA AAAAAAAA TCCACAAA CAAAAAAT ATAAAAAG AGAATTTAA 120
AATTTCAG GAATTAAT TTAATTAAT AACTAGCA TTTATTAAT AATTAATAT 180

240 TACCTACCA ACATATCTCT GCGCTCTCTC CTGCTCTCTAT ATAGAGAGCA TAGCCGATTC
240
300 TCTGCGCGGG CCCCTCTCTG ATGCTCTCTCC GGTCTCTGCTT CCGCGCTGGG TCTCTGGGGA
300
360 CCGTCCAGAG GTGAGAGTGG CTGAGTGGCC TGGCTGCTTG GTGCTGATG GTTTGCTCTC
360
420 CCGTACTCTT TTTTCTTGAG TTATCTCTCA TTGATTTTTT TCTCTGCTTT CTGCTATAC
420
480 CACCTCTGGG GAGAGAGTA ATAAACATG TATATTTTTT AAGAGAGAA AAAAAAAAA
480
504 AAAAAACTGG GCGCGCGGCC CGAA
504

15

(2) INFORMATION FOR SEQ ID NO: 52:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 777 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

60 NAAATATCTT GCGCATTTTA TTACAGAGCA GGTAAATCA TTCAATCTGG ATAGGCGATA
60
120 ACATACAGAG AATGAGACTA TCCGAGAAAT GCGAGCGAGC ATTGGAACCA ACATGATAT
120
180 CTGAGGACA GATGAGTGA TTCTGCTATT GTTAGGCTTG GAGCAATGG TCAGAAATAG
180
240 CAAAAATGG ATACAAAGG CATATATGT CACCGAGCT AATGAGATA GCTGGCGAG
240
300 TAGGAGAAAT GCGATTTTGG CTCTACACTA AGTTCTGCAA CTCTGATGA GCGTCCAAA
300
360 ACAAATTTTA GGGGAAAAA AGCGAGCTGG TTATGAAAG ATATATCTCA TTTCTATTA
360
420 AATCAATGT CATGCTCTTT AATGAGATCC TTTTATCTTC AGGAGAGAGG CATGCGCTA
420
480 AAGAAACACC AGCTCAAGAG CTTCTGATGC CAACCTAGAG GGTACCCAAA CACAAACTTA
480
540 GCAATAGAGT AGAATCTCT ATGTCTTTTG GTGAGGCCAA ACCATTTGG TTGCTACTTC
540
600 ACAGGAACAT CTTTCTACCA AGTCTTATTC ATATGATATG TCCACAGAGT CTCGATTTGT
600
660 TTGCGCACT GTGTCTATGC TCGAGATAGG CTCGAAGGTT AGTTTGTATC CTGGAACCT
660
720 ATTACAAAT GCGAGCTAAT GTGCTCTGCT CCACTTAAAA AAGGCTTGGG TCTGGCAGAG
720
777 GCGAGAGGEC CTGTGCTCCC CTGTGCGCA TTCTGCGCTC GAGGCGAAT TCTATAT

55

(2) INFORMATION FOR SEQ ID NO: 53:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 602 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

5 ATAGTACAG TGTATACCC TCCATCTTT GAGGTGGCC ATCGAACACT GCTTGTATCA
60
120 CTCTGTGCAC GGTATAAATC CATATATCCA CAAAGACCA CATCATCCA TCMACATATA
120
180 CATGCTTTGG GATGAGCAGG TCAATGATTT TGAAGGGAG TTTTGTCTTT TTTTCTTCT
180
240 CATATATCT TTAATTTGTT GTCACTATTC AAGCAACAA ACAGAAATAT TTTTGGAAA
240
300 AACCTTCAT AGCGCTTTTC TATCAAGTCC TTTAAATAT AGACTAAATA CACACATCT
300
360 GCGATTTTT TCTTACAGTG ACATATCTCT TACCTGCCAT TTATATATAG CCTGTATTT
360
420 TTCTCACTTA TATTTACTCT TCACTTGTAT TTTTATTTTA AACAGGAAA AAAAACTCA
420
480 AAAAAAGAA AATTAACCTT AGCGCTTCAT TATCTATTTA TATATATAT TATATTTGCA
480
540 CATTTTGGAA TACTGTGGAA GTTTTATCTC TTGCATATAC TTTATACCGA AGTATTTACG
540
600 CTTAAAAATA CAAATATAAA TTTTACAGG TCCGCTTTT GGTGTGGAA AGATTAATTT
600
602 CA

(2) INFORMATION FOR SEQ ID NO: 54:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1749 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

40 AGTCACTGAC TTGAGCGGCG TCGGGGGAG TCCCGGCCAG ACAGGCGGTG GGTGGGATG
60
120 CCGTACTTCA GTTTGAAAGG GTTCCGATTC CAAAGGGGTT AAAAAAGCG AACCCGATC
120
180 CCGGACACCA CTTCGCGCTT CCGTAAGAG CACACCGCG TACGCTTGG CAGCGCGGAC
180
240 CAGCTGTCCA AGTCTCTGTT GTTGGGGGAC GCGCGAGTGG GCAAGAGCTC GCTGTGTGAG
240
300 GATTATTCGC AGGACAGCTT CAGCAACAC TACAGTCCA CCGTGGGAGT GATTTTCTCT
300
360 CTGAGGCTTC TCCAGTGGTC TCACTACAG ATATGTGGGC TTCACTGTG GATATTTCCA
360
420 GCGAGGAGC GCTTCACCTC TATGACACA TTGTATATTC GGAATGCTTC TCGCTGTGTT
420
480 ATATATTTTG ACCTTACCAA TCGCACTACC TTCAGACAA CCGAGAGGTG GAAACAGAC
480
540 CTAGACACA AGCTCAGACT ACCCATGCA GAGCGGTCG CCGCTCTCT CTGCGGAC
540
600 AAGTGTATTC TGTGCGCTTG GCGATGAGC CCGGACAGA TTGACCGGTT CAGTAAAGAG
600

5 AACGCTTTCA CAGCTTGGAC AGAACATCA CTCAGGAGA ACMAAATAT TAATGAGGCT 660
ATGAGATCC TCATGAAAA GATCATGAGA AATTCAGAG AGAATATCAT GTCTTTGTCC 720
5 ACCGAGGGG ACTATCATCA TCTACAAACC AATCTCTCCA GCTGCTCTG CTCGTAGTAG 780
TGTGTGGCTT ATTTCATC CAGTTCTGAG GAGCTCTTTT AATCTCTTTC CTTTGGTTG 840
10 CCCAGCTGAC CATTTTATTA ATTCATTTTG AATTTCTGTC TGACTACTGT CCAATAGAGA 900
GGGCGCATG TCATATGAAA AAGACACTG GAAACCATGT GCATTTCTCC ATCTCTCTGA 960
TTAGCTTTTC ACATTTGCT GTCATCATGT AATGCAATTT AATGCTCTG CTCATAGATC 1020
15 TTCTCATGAG CCCCAATTT GTATTCGGA APTTTGTGAG AAGGATTAGA AATCAGCAC 1080
TCCCTTTTAG AGATCATAT TCTCAGTAC TTCTGAGCTT ATTTTTCTAT TTGATATCA 1140
TTGATATCAT GACTTCATAT TGAAGAGAA ATGATATCAA ATGTCATTTT CCAAAATTTCT 1200
20 TGTAGGCGCT TGTTCAGAT TCTTTCTGTC TTGGAATGTA AACATCTGAT TCTGATGTC 1260
AAGAGAGGG GTCTGGGAT CTGTGAAAT TTGGCTACTA GAAGTGTGCC AGAATGACT 1320
25 GTATTTTGA AACTCTTAC CTCATATTA APTTTCTCTT GTCTTGCAAT CAGATATGAT 1380
CAGTTTTTTT GGGCGGAT GTGGCTCAT GCTGATATC CCAGCATTTG GGGAGGCCAA 1440
30 GCGAGCGGA TCATCATGAG CCAGGATTC GAGACAAAC TGGTCAGAT GCGAAAGCCC 1500
CCTCTCTACT AAATATGAAA AATTAAGCA GGGGTGATGG CAGTGTCTG TAATCCGAGC 1560
TACTCTGAG ACTGATGAG GAGATGCTT TGAGCTGAG AGCGAGAGGT TCGATGTGAC 1620
35 CAGATATGAT CAGGCTGACT TCAGCTTGGG TGACAGAAA GGAATCTCTC TCMAAAMAAA 1680
AAAAAAAAA AAATCTGAG GGGGGGGCG GTACCAAT CCGCTGATA GTGATCTGAT 1740
40 AACTGAAA 1749

(2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1896 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

55 AAGAGATGG GCTCTTATTT TTCTGAAA ACCAATTTGG AGTTACTACT TTTTCCATTA 60
CATTAATTT CTATAGTGA ACTCATATTT GTCCATGAT GCTTCATGAG GACTCATGCC 120
CCTCTGTCT ACTGCTTCA AATAGACAT GTAGCTTCA CCCCCGTGCT TTGTGTCTAT 180
60 GGTGGGCTG TGTATATGG AAATATGCA GGTGTGTCAG GTTGGGAGAC AAGATATTT 240

5 TTTATATGCT AGAGCTTTA AAAACCCAG CAGAAATTA TTCAATATTT GTTTATTCOC 300
TTTTTTTGA CAGATTTGTT AATTTAATG AATTTAAGG GAATCTGGA GTACTAGAC 360
GTTTATTA AAAAGAAAAA TGTCTCTAAA TGTCTCTAAA TCACAGAGG AGAAATATAC 420
TTTGTTCCTT GATCTGTGAG AATGCAAT AACTGTGGC GAGCTGTAT TATTTATAT 480
10 AATATATAG TACGAAGTTA ATACTGCTT CTCTGTGTC CAGGCAAT ATTACAGCTT 540
CTTTTGAGC GTAAATATCA GAATGATCC TCTTCCAGG GAAATGACA GAAGCTATAT 600
GTTTACAT GTTTGATTT GTGATCTGAA ATACACAAA ATTAAMAAA TGAATTTCT 660
15 AATTTTCAA CTAGAGGAG AGAATCTGT GAAATATTC TTTTATTTTC TTTTATTTT 720
CTTAAGAGAG GCGAGCCAG GTAGTAACT AAAATATGTC CCGAGGATA TGAAGTTGT 780
20 CTTAGAGGT TAAGAGAC ACCTGTGAC TGTATGCTT TGGGCTGAG TGGGCAAGGA 840
GCTCAACTG ACCGTGCAAT GTTGTCTGA CTATATAGA CAGGCAATC ATTGTATTC 900
TTGACAGGG TCTGAGACC TGAAGATG TTATATAGA GAAGAACT CTTTCTGAA 960
TATTCATGTA TAAGAGACA AATTAATTT TCTGAACTTC TGCATCTG AGAATCTGC 1020
AAGATTTTA CAGTATTTT AGCTGTGTC AGCATTTTC CATGAGACT TTCTATCAT 1080
30 TGAATCTTTA GTTCAGAGT TCCATGAT TGTGAGCAG AATATATCT CTTTATGCT 1140
TGGGATCCA GCTGAGGCA ATCTCTGCA TTTTATACC GGTGTATGTA CAGATATCAT 1200
TCTTGTGTA TCCATGACT TGAAMAGTT TGGGAAGCTC TTATGCAATA TCAGCTAAA 1260
35 GGAATGAAA TCAGAGTGA TTAGGTTGT CATTCATGTA TTCTCTGAA GCGAGACCC 1320
AAGGAAATA TAGTCTCAT CTTTACTAT CACTCTAAA TTTAATGTA ATTCTATCA 1380
40 TGTATATG TGTTTTCTTT ATATTTTAT AAAATATAT CATGCGAGT TTAACTTCA 1440
CTTCATGCT ATCGATGTC TGGGCTCCA TCGAGAACT TCGAAGAAA ACAGGAGGA 1500
ATGCAATTC ATATGAGCC AGATCATAT TTCTGCAAC TGAATATAT ATTCTATAT 1560
45 TGTCTTGA AGTCTGAT TCTTACTTT TCTTGTGTC ATATATAT AGTGTGCTATC 1620
ACTGATAT GTGAGTGAC TAGAGATGC ATATATCAT GTTAAGTACA GTTAAGGATA 1680
50 AACTTTGA GGTATATAT TTTAAMAAA CTTTCTTGA ACTCTGCGC CAATATGCT 1740
GAAGACCCCT CTTCTTACTT AAAATATCC AAATATAGC CAGGGGGGTC GATGCGTGG 1800
GTGCTTTA ATCTTCACT ACTTGGGA GGGCTTGAAG CAGGAGGA ACTGCTCTG 1860
55 AATCCCGGGG NCGGCGAGTA GGTTCGAG TTGATG 1896

(2) INFORMATION FOR SEQ ID NO: 56:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1753 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

10 TCTTTTAAAT ATACACATTT GTGGGCTCA CACATATAT GAATAGTAC CCTCTAATA 60
AGAGAAAAA AATATGAGC GTTCACACT AGAGACATAT TGTCTATTA AGCAATAGT 120
15 TATTTCACTA GAAAAATTT AATATCAGG ACTATTACAT ACTTCATTAC TAGGAGTTC 180
TTTTTAATAT GACACTTAAA ACATCTACTG AATCTTTGAT CCACATACA CCTTTTAT 240
TTTCTCTAAA CACTTTGAAA GCTTAAGCTT CTGACATCA TGTGCAAGT GTGATGGCA 300
GTAAATATAC AGAGAGATG TTATGTAGCA ATTAAGGCT GTTTCAGCT TTAGAGACA 360
CCTGGCTGT ATGATCTCT GGGGCGAGG TGGCATATG TTTTTCAAA ATATAGCAT 420
25 ATGTACATG TTTCATATT TTTTGTCTG TTCAATTTT GACAGCGAG TTGACCATC 480
ATAGAAATTA TTACTTTCTT TCAATGTAT TTTGTTTAC TGGCTTAAA GCTTCTCAG 540
AATATCATG GCGAGCGAG CATACAGTT TCCTCTTAA TAGAGATGAA ATTATTTTG 600
30 TATCTACTGA TACAGATAC TGGTCAATG GAAAAAAT CATTTATCC GTCTTTTAA 660
TTATGTATTA AATATATAT TTATGTCTG GATATTTGA GACAGCTCT GAGAGACAA 720
35 GTTTTCCAT TACTCTTCT GACCAATAG CTGCGACCG TTGCTTCTC TTGGGAGAA 780
GGAGAGGCTG TGTGACATG GCTACCATC TTCCATATAC CAATATGCA TAGCTAAT 840
40 AAGTATTTA TTTTATGCTT CAGTATATA TTATTTAAT TTTTAGTAAA TGGCTATCT 900
TTGCTCATT AGGAGAGAA GCAATCATGA GAGATTCAG GATAGTTTG TTTAATTTCT 960
TCCAGATTAC ATGTTTTAC AGTGGCTGC TATTTAGGAA AGGTATTTCT CTATACACT 1020
45 TGTTTTACC TTTCAGACA TTGAGGAAA TTATCATAG GTTTGTTGAG ATACGACTT 1080
GATGGCTG TTATATAGT TTGCTTCAA AGTGGCTAC TGAAGGCC CTAAAGCTG 1140
50 TAGAATTTA AGGATTTCA AATTTTCT ATCTCTTCT TTAACTTACC ACGAATAG 1200
GATTTGATA GCAATGATG GTATGATGA GAAGTTTGA CCAATTTGT TTTTGTGTG 1260
TTGTGTGTG TTGATTTG AATTTTCT TATTTCTTT AGAATGTTT ATGTATGAT 1320
55 GTGAGATGC TAGGAGCT ATGCTCAT ATTCATGTA AGTCTGCTT CACCTTTAC 1380
AGATTTGAG ATGCTGACT GATGATAT ATTTCTTAG TAGATATG TTAAATATAC 1440
AATGATTTT TAAAGATG ATGCAATCT GTATTATG TGGTGTCT GTCTTATG 1500

5 GGAGCAATT AATCAATTT CATATGATT TTTCAGTGT TGAATCTAC ACATCTACT 1560
TTGAAATTT CCTTCATCC TGAATACGA ATAGAGAGG CCAATATAT TCCCTCTTA 1620
10 TCCCTGAGT TTCACTACT TTATGTTAAA AGTTGTGTAT AATGTTAAA ATCTGTGAAA 1680
GATTAAGAG TCGATTAAA TTAAAAAAA AAAAAAATA AAAAAAATA AAAAAAATA 1740
AAAAAAAAA GGG 1753

(2) INFORMATION FOR SEQ ID NO: 57:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1220 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

25 GCGAATTTA CTGACGCCG GGTGTTGTG TGTGGGAGG GAGAAAGAT TTGTAAACC 60
CGAGCGAG TTCTGCTTAC CCGAGGCCG TGTGTGCGG AGACCTCGG GTGAAGCCAC 120
CGTATCTG TCTGACGAG AGCGAAMACC TTCACTGTAG GACTTGGGG ATTAAGAGA 180
30 AGCTGATAT ATTAACTCA AATGATTTG AGAGATAGC AGTGAGATTC ACTTCAGAT 240
GAAATGACA ACACATCTCA AGAATCTCA AATATCATAC TGTCAAGAC AGGTGTCTC 300
35 AATGATTTA CTGAGTTTC TCTTTGAGG TCAGAGATTT GCTGATATC ATACTCCAA 360
AGACTGCGA ATGAGGAGG AATATGAT TGAATTTTAT CAGGACAAA CCGGGGCTCA 420
TTCAAGATT TAGATATCT TTTTATTTT TTCTTTTTC CTCAATCTT TTTTATTTT 480
40 AAAAAATTT CTTTGTGAT GTGTGTTCA AAGCGAATTT GAAAATCTGC ACCCATCTC 540
TTTGAACAT CTGTATTTT GAATCTAGT GCTCATATT CATTTATGTT TGTTTTCAT 600
GTCTGATTT TTGCTATCA AGCTCATGC CCTTCATAT TACCTCTCC TTTTAAAAA 660
45 TTACTGTGC ACAGAGAT CACCTTTTTC AGGACATGCG ATTTTCAGC TTGTGTGAT 720
AATATATC GACATGCA AGTGTGATA ATGACTTTTC AATGGCTCT GATGTTCTAG 780
50 CATGTGATA CTCACTCT GCACTGAC TTTCAAGTGG AGATGAAAT TTTTCAGCA 840
ACTGACTGT GAAAAATCA CTTTCTCTA ACTGAGCTT ACTTTTAAA TTTCAGGGCT 900
TGCACAAA GAAGCAAT ATCAGTTGA AGTCAGATG ACAGATAGG TGAGATATAT 960
55 GATTAATCTT AAGATGCTT TCACTAGA AAGGCAATTT TAAGATTTT TAAAAATCTT 1020
GTCAAGAT CCGAAGAG TTCTATTTT CATTAAGCAT TAATAAGCT ATACATCAG 1080
60 AATGATAT ACAGACAC TGTCTTTT GATTTATTT GTACTTTTG GCTTGGATA 1140

TGGGTTTAA ATGACATG TCCTACACAG CTTCTATAAA ATAAACATA TTCTTAAAA
TCAAAAAA AAAAAAAAAA

5

(2) INFORMATION FOR SEQ ID NO: 58:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

TCGCGCTGCG AGACACAGCA TCCTACTGAC GTGGGTGACC TCTGTGACCA TCCTACTGCT 60
CAAGCTGCCC TCATTTCTG GTGAGGCCA TCAGGGAGCC ACAGCGCTCG GAGAGTGGT 120
CGGATCTGTC GCGATGCGG AATCTGTGA GAGAGAGGAC CCGCGATGTA GAGACACTAC 180
CGAGCGACCA AGAGTACGA TTCTGTGCA GAGCTTCTTC AATAGGGGCC ATGCTGCTCC 240
CGCAGGGGCT CTTGGGCCCC GCGAGACGA GGCAGGTGCGC AGGCTGGGTC GTCTGATG 300
CGCTTCTCAT GAGCTCAACC GCGAGCTGCT GAGCATGGGC TTTCGCGAGT GGCATCTCG 360
CAACCTGCT GTGAGAGCGG TGACCTGCA CTCTGCTGTC TTCTGTGCTCA TGATGCTTGG 420
TGTCTGTGGC CTCCTGCTGG TTGGCTTGT CTACTGTGTC TCCCACTGA GTCCAGGCTG 480
ACCTCTGAGG GCTGATAGGG GTGGGTTTGT TGAGAGGGAC TTGCTGGGCC TTGTTGTGAG 540
AGCAGGCATA TTTCGAGGG ATCTGTGCT GTCTTGAGGG TATGATCAGA GAGGGACCA 600
CAGGTGTGTC TTTCGCTTTT GTGTTAGGG TGAGGGGAG GAGAGCTTGA GTCCACAGAT 660
TTCCCAAGT GTGGGTGGCT CGGTGCTTC CGAGATACT TTTAGGTGCT ATCGGGCTCG 720
CATTAAGTGG CAGAAATCA GAGCAGAAA CGGATGCGCT TCCGATTTCT CTCATCTCTT 780
TTATGCGGAG AGATCTCAG CTGGATGCA ACATGTTCC ATGCTGTGCG AGACATGCC 840
GAGCTCTCT CTGCTTAGGG AGCGAGATT GGGCTTAGGG CAGGTGGAAT AATTTCCAGA 900
CTTTTTAGC ACTGTTTGTG TTTTATAGT ATATTTTAT TGCTACTTTT ATTGTTTAG 960
ACAGTGGTA GTGGATCTCT ATTATGTG ACCTTTTCAA TAAATAGATT TAAGTAAAA 1020
AAAAAAAAA AATCTCGAG GGGGGGCC 1049

55

(2) INFORMATION FOR SEQ ID NO: 59:

(1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAGAGATG TGAAGTACA GTTCCGAT GTCTGTGCG ATGGAAGCA CAGGCTGAG 60
GAGCTGATT GTAAAGGAC TAAATGCTTC TCGGTGAGA ATCAGATG GGGACTCTT 120
AGCTTCAT CTCTTTCTT TCAAGTCTG GACATCTGA GGTGAAGTC GGCATCTCA 180
GTCAAGTAT GAGTGGAAA GTGAAGTGA CAGGGGAAA CAGAGGAAA TTCTGTCTTC 240
TAGGCAATG GGCAGAGGG GCAAGCTG GCACTCTAT CCACTCTAT CCACTAGTG CTGAGGCCC 300
CTGTGCTTA CTTTCTTGA GAACTGTG AGATGCCAA TTTTAGAG CTGCTGAGA 360
GTGACTTTC CTCTACCTTC CGGCTGCTA CAGTCTTTC TTATGAGCA TACGCTGAT 420
ACTTAGTGA AGCGCGGAT CTCTCTGAC TACAGAGC TGAAGAGAT AACCTTGAC 480
ACCTCTGAT TGTACCTG CTCTTAAAG TAAATGTAT CCAATATCA GTTGTGCTG 540
AGGCTCTTC CTTCGATAT GTGGGAGC TGAAGACT TGTATATG AGTGTGTG 600
CTGAGTGT TGTGTGCTC GTGAGCAG CCAACAGG GCTGAGATT GACTAGCA 660
TGGGCGGGA CATCGAGCC CAGGACTCA GTGCTATTC CGGAGCTTC AATGAAAC 720
ATTAAAGTGA GAGGGGAG AGGAGGACA GGCATCTC AGGAGCTGA GCAAGCTG 780
ACTGAGCTA GCGAACAG TCTGTGACC AACAGGCC ASCAGCAG AACGCTCAA 840
AGGCGAAGG GCTCGAGGG AAGCGCAGA TTTGTGCAA GTGAAATTA AAGTACTC 900
GTTTCTCCC TGCGATGTC GGTGCGAG TGCCTGCTG CTCTTAGCA GTCTCTAG 960
AGGCTCTCT GCGCTGCCC AGCTGATAT CTGAGTTCA TGAAGCTTCA CTTCCCTTAA 1020
CGCAAAAT AGATCAGAC TTCTCTAGG AGGAGTCAA TGTAGTCTAT GTTTTCTT 1080
GTACTTCTG TTTTCTGTA CTCTATGT TCAATGCTC CCGCTGACA TGTCTCTCC 1140
CTTGTCTCT TAAAGCTCA GATCTGTC CTGTCAATTA CATGTCTATG AATGATGAG 1200
TAGGCTGAT GGGGTGCT CTGTCTGAG CATTAACCA AGGCTTTT TCTGCGACC 1260
CATCCCTGA TGCCTGATC CCACTCTTA TACCTTACC CTGACTTAT GAGAGCTTC 1320
TGAAGACA TAGGGCCCC ACCTTTACT ACAGCTGAG AATCTGGA GCGATCTCT 1380
CATCCAGGA GTACTGAC ATGTCTATC CTGATCTG TGAATCTG TGAJCTTCA GTATTTCT 1440
TTCTCTCT TGCGGCTTGT GTCTGAGAA TGTGCTGCT TCAAGCCAG AGCTTAGAA 1500
TGCGAGGCT TTCTTACT GTTGAAGT GATCTTTCT TGGGCTTGC CATCTGAGA 1560
AGCTTGAAG CAATCTGGA AGGTTTAT CTCTTTTGT GAGTTGCTG GCGAAGGAA 1620

TTGGTTTATA ATGCAGTGTG TCTGTACAGG CTTCATTTAA ATAAACATTA TTTCTTAAAA 1200
TCMAAAAAA AAAAAAAAAA 1220

5

(2) INFORMATION FOR SEQ ID NO: 58:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

TCGGCCCTCC AGACACAGCA TCTACTGACG GTGGGTACCC TCTGTGAACTA TCACTGACTG 60
CAGGCTCTCC TCATTTCTG GTCCAGCCCA TCAGGGACCC ACAGCGCCCTG GGAAGATGGT 120
GGGATCTTG GCGAATGGGG AATTCGTGCA GGACAGCAGC CCCCAGATGA GACACACTAC 180
CAGCGACCA AGAGGTAGCA TTCTCTGACA GAGCTTCTTC AATAAGCGGCC ATGCTGCTCC 240
CCGAGGGGT CCTGGCCCC GCGACAGCA GCGAGGTGCG AGGCTGGGTTG TGTCTGATGC 300
CCCTTCAT GACTCTACC GCGAGCTGGT GACATATGGC TTTCCGCAAT GGCATCTCG 360
CAACATGCT GTGAGCGGG TGAAGCTCAT CCTGCTCTC TTCTCTGCTA TGAATGCTTG 420
TTTCTGTGCG CTCTCTCTCG TTGGCTTTGT CTAGCTGGTG TCCCACTGTA GTCAAGCGGT 480
ACCTCTCAGG GCTGATAGGG GTGGGTTTGT TGAGAGGAGC TTCTTGCGCC TTGCTGTGAG 540
AGCAGGCAAT TTTCGAGGG ATCTGGTGGT GCTTTGAGG TATGATCAG GAGGGGACCA 600
CAGGTGTGTC TTTCGCCCTT GTTTTANCG TGAGGCAGAG GGAGACTTGA GTCCAGCAT 660
TTCCGAAAT GTGGGTGGGT CGGTGTGTC CCGAGTACT TTTAGTGGT ATGGGGCTTG 720
CATTAAGTGG CACAAATCA GAGCAGAAA CCGATGGCT TCCCATTTCT CTCATTCCT 780
TTATGCCAG AGATCTCAG CTGATGCCA ACATGTTCCG ATGCTGTGAG AAGACATGCC 840
GAGCTCTCT CTGCTTAGGG AGCAGACTT GCGCTTAGGG CAGGTGAGAA AATTTCCAGA 900
CTTTTTTAC ACTGTTTTTG TTTTATGGT ATATTTTAT TGGCTACTTT ATGTTTAGG 960
ACAGTGGTA GTGCAATCT ATTATTTG ACCTTTTCA TAAATAGATT TAAGTAAAAA 1020
AAAAAAAAA AAAAAACGAG GGGGGGCC 1049

55

(2) INFORMATION FOR SEQ ID NO: 59:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAGAGATG TGAAGTATA GGTCCCGCAT GCGTGTGCG ATGGGAAGCA CAAAGCTGAG 60
GGACTGGATT GTAAAGGCAC TAAGTGTTC TCGGTGAGA ATCAGACATG GGGAGCTCT 120
AGCTTCATAT CTTCTTTCT TCGAGSTCTG GACATCTCTA GCCCAAGTCC CCGACACTCA 180
GTCCATGAT GAGTCCGAA GTGAGGTGA CAGGCGAGAA CAGGAGCAA TTTCTGCTCC 240
TAGCCAGTTC GCGCAGGGG GCGAGGCTGG CCAACTCAT CATTAGGTC GTGAGGCCC 300
CTGGTCTCTA CTTTCTGCA GAACTGCTGG ACATGCCAA TTTTAGAGAG CTGCTGAGA 360
GTGACTTTGG CTCTACTTTC GCGTGTCTCA GAGTGTTC TTATGGGACA TACGCTGACT 420
ACTTAGCTGA AGCCCGCAT CTCTCTCAC TAAAGAGGC TCAGAGAT AAGCTTGCAC 480
ACCTCTGAT TGTCAACCTG GCTGTGAGG TAAATGTAT CCGATATGCA GTTCTGTGCG 540
AGGCTCTTGC CTTGCTTAT GTCCGCGAG TCGAAGACT TGTGATTGAG GCTGTGTATG 600
CTGAGCTGCT TGTGTGCTCC CTGAGCAGC GACACAGCG GCTGAGGTT GACTTACAGA 660
TGGGGCGGGA CATTGAGGCG CAGGACTCTA GTCCCATTC GCGAACCTK AAAAAAAC 720
ATTAAAGTTA GAGGGCAGC AGCAGCCGA GCGACTCTC AGGACCTGA GCAACACTCG 780
ACTGAGCTGA GCGAACGCG TCTTGGCAC AACAGCCC ASSCAGCAG AAGCCTCA 840
AGGCGAGGG GCTCCGAGG AAGCCAGGA TTTGTGCAA GTCCAAATGA AAGTACTCTC 900
GTTTCTTCCC TGGGATGTC GGTTCGCGC TGCCTGCCG CCTCTTAGCA GTCTCCAGAG 960
AGGCTTCTGT GCGCTTGGC AGCTGATAT CTAGGTTCA TGAACCTTCA CTTCCCTTAA 1020
CCCAAGCAT AATCAACAC TTCTCTAGG AGGAGKAAA TGTAGGTCAAT GTTTTCTTTC 1080
GTACTTTCTG TTTTCTTCTA CTTCATGCT TCCATGCTC CCGCTGCCA TGCCTCTCTCC 1140
CTTGTTTCTT TAGAGCTCA CACTCTGTC CTGTCTATTA CATCTCATTC AGTAGTGGG 1200
TAGCCCTCAT GGGGTGCTT CTCTCTGAG CATTAACCAAC AGGGTTTTTT TCTGCCACCC 1260
CATCCCTGCA TGCCTGATCC CAGTTCTTA TACCTTACC CTGAGCTATT GAGCAGCTTC 1320
TCAGAGCCA TAGGGCCCC ACCTTACTC ACACCTGAG AATCTGGGA GCGAGTCTCC 1380
CATCCGAGA GTCACTGAC ATTCTCATCC TAGAATCTCG TCNCACTACA GTCATTTCTT 1440
TTCTCTCTTC TGGGCTTGG GTCTTGGAA TGTGTGCTCT TCAACTCCAG AGCTTAAAGA 1500
TGGCAGCGGT TTCTTAACAT GTTAGAGAT GATCTTTCT TGGGCTTGGC CATCTGGGA 1560
AGCTTGATGG CATTCTGGA AGGTTTATAT CTCTTTTGT GAGTTTGGTG GCGAGGGA 1620

60

GGGTATATAG ATTGTATATTA AAAAAAAAAAAG GTATATATATCC ATATATATATAT ATATATATATG 1680
ACCGAGAAAT AAATCTATATCA GAATCTATATC TACAAATATTA AAAAAAAAAA AAAAAAAAAA 1740
AGGATATCCA TTTCATGCTT ATCGATACCG TCACCC 1776

10 (2) INFORMATION FOR SEQ ID NO: 60:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ACAGTAAAT AATATATATTA TAAATTAAT TAAATATAAA ATCTGAGCTA ATCTGATTTA 60
ATTGAGAGAT TTCAATATGA ACCAGGATTT TCTGGCTTCC CAGGAGAGAT CAGAGAGGCT 120
AGCTAGCAAC ACTGGCTGCG TTGGCTACTT TCTTTGGAC AACTGAAAT CTAGCTCCCT 180
TTTTTTT TTTTGGCCC ACTTCATCCA TTCACTATGAC CTGGCTGGCC TCTGAGGTTA 240
AGTGAGTATG CAACAAAT GTACGACGAG TTTTGTCCCT GAACTAGCTG GTTTGAGGTC 300
GAGCTCTGCC ACTTGTGACG ATGACCTGGT GCGAAATTC NCGACGAGAT TTTTTTTTTT 360
TTTTTCAATG CTCAGTCCC CCTATGTGAG AATCTGCCCC CCCCCTGGCA CAGATATTTT 420
ACCTGGCCCC CCGAGTCCC TCA 443

40 (2) INFORMATION FOR SEQ ID NO: 61:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2888 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTATGTGTTT CATTATCAC CAGGCCAAC AGAATTTATA TGACCTGGAT GAAGATGATG 60
ATGGTATAGC TTCCGTTCCT ACTAATACGA TGAAGTTTGC AGCTTCAGGC GNCCTTCTCC 120
ACCAGTGGC TGGGTATAGC AGTTTCAGGC TTTCATATGTC CAGGCCCTC CTCCTACCCA 180
AAGTGTCTCA GAATATGCA TACAGAGTC CTGCTCTCCC TTCTCTTATT CGAGCAATAG 240
CCTTGACAAA CATGTGCCC ACACTGTGTA ACAGGAGAA ACCCCCCAAA GAGCTGCCAG 300
CTCTGAGCC TGTTCAGC CCATTTGAG GCACCAATG GACTGTGAT AATCTGALCC 360

CTCAGTGCAC TGAAGAGAC ATTGTGAGC TTTTCTGTGT GTGTGGGGCC CTCAGCCAG 420
CTCAGTGTCT CATTCTGGG GTAGGGGAGG TGGTGTGTGT GAAAAAGGAC GATCCCATCA 480
CGCATATTA GAATATACAC AACGGGTGTC TGGAGGGGCA GCGGATGAG TCGACCTTTC 540
ACATGATGG GAATTTATC ACCTGAGAC AGCCCATCTT GCTGGGGCTG AGTGAAGACC 600
CATCATGAA AAGAGAGAC GAGCTGGCTC GCAAGGTGAA CTCGTGGCTCC TCTTCGACCC 660
CCCCCTGTA AGTGAAGCTT GACACATCC TGAAGGCACT CTGAAATGTC TGAAGGGGCT 720
CTATGACCA CGAGCCCAACA GAATTCAGAA TCAAGCTTTG AGCAGGGGAG TGAAGGAGCC 780
AGAAATGGGG CGAGAGAGG GTGGCTCTGT TTCCCCAGG CAAAGCTTAT GACCAATGGG 840
CCATGGCACT GGAGAGCCCT GATTTGGGA AGGGTTGCC GGAATTAAGA GCTTCTCTAC 900
TGAATGGGAC CGGCTTTCTT GTTGTGTGTT CTGCGCTGTC CTCCTCTCTC TACGTTAAG 960
TTTTCTGATG TATGTTCCTT CATCTCATG CCAAGGTAGG CTGTGTGTTTT TCAGTGTGTG 1020
CTTCCCCGAG CTCAGCCCC AAGCTCAATTT CTATCTGGA AATGTATAC TGAATTCCTT 1080
GGTGGGCTTT CTGTGGGCC CATGGGATGC AGCGTGGGG CTGTCTGAGG GAGCTCTCTT 1140
TTTTCAGGGG CGAGAGGCTT GCTTTTCTTT TGTGTGATTT AAGCTTTTCA AICANATGAG 1200
GGGATGAGA GCGCTGTGTT CCTGAGCGGA GCGAGGTGG CTTGAGAGCT GTGCGCTCC 1260
TCTGTCTTCT CAGTGAAGGT GCTTGGGTTG GGAAGAGTTC TGAAGGCTCT TATGCTCTCC 1320
CGAGTGGCT CAGGCTCAC TAAATGGAG GCGAGATGTA GCTTCAGCC CTGGGAGAG 1380
TCTATCTTAG TTCTTGGCTT GGAATCCCC GTCTCTCTCT CCAAGAGGA GTTCTCCAGA 1440
GTTCAAGCTT GCTTTTCTT TGAATGTGTC TGAATCCCC AGCCAGGCTC TCTTTCCCTT 1500
CTGGGTGTCT TTGCTGGAG GGGGCTGTGT TGTGAAGCTT CCGGTTCTC ACCTGGGCTG 1560
GCACTTAACC ACAAGCTGCT TTTGTGAGC GCGAGCTCT CTTCAGGTTG GCGCTTTGAA 1620
AGGCTCAGCC TCCCATTTG CAGTGTCTTG GTTTGGAAGT TATTTGAATG GAAGAGTCTA 1680
GTTTGTCTCT GGTCTGCTAT TTCTGGCTC AGTTGTCTAC AGGAGCTGTC AGTGTGATGC 1740
CTGAGGCAAT ATATTCAGCT GCGACCAAG GGCAGCTGTT GTTCCACTT ATGTGATGTA 1800
CGCCATCAT CCATGAGCAG AGGATTAATTT TCTGTGCTTG CGAGAGGAG AGGATGCTAG 1860
GGAGCAGGCG AGCTTACCA GCGAGGTGT TTCCCAACA TAGGGGCAGA CAATTTGGAC 1920
GAATCTTAG AGCCAGGCA GTTCCCTGAT GACAGGCA GTTGTCTAC TGAATGTCTC 1980
CCTGCTGTTT GCGAGTGAAG AGAATCAGG CTGGCAGAG TGGAGGCAAT TGGGGAGCAC 2040
GTTTCTGGA GCTCTGAAA ATCAGTACCA AGTGTGGA AGGCAATG CCGAGAGTAC 2100
TGAAGGCTC CCAAGTTTG CTTCAGGCTA GCGCAGTGA AAAAAACAGA GACTCATGTT 2160

2220
2280
2340
2400
2460
2520
2580
2640
2700
2760
2820
2880
2888

5
10
15
20
25

TOCAGGGGTC AGTCTGTCTGAG CAGAGAGGGA CCGAGAGTTT GAGCCGAGCT TCGAGTGTGCA
GGCTCTGAGG GTGCGCAGGA CGGGAGAGTC CAGAGAGGGG GCTGTGTGCT GCTCCAGCTTG
CACTTCTTTTAA AGAGATGCTG CTTTATATTC TCTTACGCTT TTCAAGTGGG TCGAGAGCTTC
TCTTTAGGAG CTGAGAGACA TTCTCTCCAG ACTTTTCTCT TCTCTGCCCA AGAGAGGAGTC
CAGAGGGCAG TAGGAGCTGG TTTTTCAGCT ACTGTGGAGCC GGGGGCTGAC TCGTTTGCAGT
GTCTTTAGGG TAGGAGCTGT AAATATCTTC CCTGCATGGC TTATATCTCTC GTCTCATGCC
AAGAGCAGGTA TCTTCTGTGT GTGACAGAT TTCAATTGAT CCGAGCTGAG CAGAGTGGCC
ATCTGGAGCT GGTGCTATAG GTGAGCATCT GTTACATTTGA GGGGAGCTGT TTGCTCTCTC
CACTCTATTA CCACTCATCT TCGAGAGGCG GAGAGAGAGG GTGTTGGGCT AGTCTCTGTGT
CCTCTCTGAC TTCCGATGCC TCTATGTTAC CCACTCTGTGT CTCTCTGTGCA GAGAGAGAGG
AAGGGGCAAT AAGAGATGAA GGTGTATAT GTTATCTACTA TCCATTTCTG ATTTAGAGCTT
TGTATATCTT AAAAAAAAAA AAAAAAAAAA CCGAGGGGGG CCGGAGAGCC AAATGCGGCK
AAGTAGAG

(2) INFORMATION FOR SEQ ID NO: 62:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1851 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

40
45
50
55
60

CACTAGATATA ATTATATAT ATTACCTAT CTGATTTCTT TTCAATATAT AGGTGTCTTA
GTTCCTATAG AGGTTTGGC ACTTCACTCT GCACTGTTTC CCAACTTTTG GACTGAGCTA
TCCGAGCTC AGTCTGCTAT GTCAAAAAC TGCATCAGC TTTTGTGTGA AGATCTGTGT
TTCCGAGAT ATATATATAG TATCTATAG GATCAAGAA CTTTATTTAA CAGACAGAT
GTCTAGACT TCAATGACA TTCTTTCTTA AGGTTTCAAA GTCAAGTGT TTCTGAGCA
AACTGTGCA ATTGATGAG CACTCTTAT ATAACTTTGA TAGGAGTA TCGAGAGCTA
CACTCTGAT TTCTCAAGCG AGTTGAAAT TCCAAAGCA GTCTTCTTT AAATGGGAC
CTGAGGGCAC TCGTTTCTCT CTTCTGATTA CAGACTGCA AAGAGTTAA CCGAGCTTA
ATTCCAGCT TCGAGAGCT TTATGAGAA TCGAGAGCT GTCTGCAGCA GAGAGCTCA
CTCGAGAGC AAGAGGCAA AAGAGAGAA ACTATAGAT ATGAGAGAG AACTCCCAT

660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
1380
1440
1500
1560
1620
1680
1740
1800
1851

5
10
15
20
25
30
35
40
45
50
55
60

AAGAGGGGCG GTTTTAGGAG TCAATGAGAG CACACTGTAG ACAGCTGCAT CAGTGCAGTC
AAGAGGAA CCAAGGAGGT CCTGAGCCCA AGAGAGACTT CTGAGAGTGA GACGAGAGAC
TCTCTATTTA TTGATTCAGG AACTGAGCAA GATCTTCTCT CCGCTGAAA TAGTTCTGT
AAGATATACC GAATGAGAGT TCGATCTTCC TTTTTCAGAG ACATGTGAAA TATCGAGTCA
CAGCTTCAG AAGAGAGCTC CAGCATGCTT AGATATGAGC ATTGTAAAGA ATTTAAAGAC
CTCCACTGTT CCAAGGATTC TACCTTAGCC GAGGAGAT CTGAGTTGCC TTCTACTTCT
ATCTCTCAG TTCTCTCTGA CTTAGCTGAC TTGAGAGCT GTCATGGCCA AGCTTTGCC
TCCGAGACC CTGAGTTGCG TTATCTCTC AGTTGTGGCC ATTTCAGAGG ACTCTTTAGT
CATATGAGC AAGATGAT TTATAGATACC CTGTGTAGGA CCAATTGATC TACATTCAT
GTCTGACAA GCATATCTCG CAAAGGAAAC CAACTCTCTT CTTCAGATTA GGTGTAGCAT
GTCTACTTTT AAGTGGCTCA CCGGAGGCC CAACTCTCTT TGTATAGTT TTGCTTATTT
GTTTTGTGCT TTCACTTTGT CCACTGCTCT CTCTTCAAT GCGAGATAG ATTTATAGCC
TTATATCTTG GTTCGAGCA ACTCGAGATG AAAAAACTT GCACTTTCAG TATACTTCT
AAGGCGCAAT CAGATATGCG ATATGTTTAA TGTATATAG AGTTCACTTT AGTGGCTTTC
ATTATATAG CTGTCTCGCG AAGAGAGGG TTGCTTAGCC CTGTACAAAG TATTTTAAAC
TTACAGCAT TTACTGTGT ATCATATGCT GTCTCTGTG CCACTTTTGT ACTTATAGA
GCGAGATTC CTGAGTGC TTGTGTTCTT ATATCAAAA TTAGTTTAC TTGTATAGCG
AACAAGCA AAGATTTGA TTCTGTAAAG AATCTCTTT AGCTGTGGCC TCGCATATA
TAAATGCTGC TTATTTTAC AGATACCTG TCGAGGAAT AAGGAGACT TGTATTAAGA
ATATTTTCT TATTTTAT GACATGACT ATGATTTCT ATTCTGTGCA CTTAATTAAG
CTGATTTGCA TCACTTTTAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A

(2) INFORMATION FOR SEQ ID NO: 63:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3542 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

60
120
180

TCATGCTG ATGAGCTGT TCGTGGCAG GCGAGCTCT TCGTCTGTA TGACTTTTC
AAGAGATTC CCACTGCTCT GAAAGATTC AATAGGTAT CCGTATCAT TGACCTCTCT
GGCAGGCC CAGATTCAT TATGATGAA TATAGGATC GTAGATCAC AGGAGCAGC

5 TTCTGTGATG AGCGCTTCAG AAGAGACTTA GAGATGTCAC TGAGATTCGG TAAACCCCTT 240
CTGTGTCAGG ATGTGGAAAG CTAGATTCAC GTTTTGAAAC CGGTGTGAGT 300
CGCGAAGCAG GCGGAGAGGT CTTGATATCT CTGCGGAGACC AGGAGATAGA CTTGTGCGCA 360
TGTTTGTCTA TTTTCTGTTC CACCGGGAT CCAACTGTGG AGTTCCGACC AGATCTCTGT 420
TCCCGGGTTA CTTTGTGTAA CTTGACATTT ACCGTATGCA GTTTTACAGG CAGATGTCTA 480
AATGAAGTAC TTAAAGCAGA AAGAGCTGAT GTGAGCGAGA AAGGATCTGA TTTTCTTAAA 540
CTTCAGGGGG AATTTCAGCT CCGTTTGGCT CAGCTGAAA AATCTCTACT AGAAGCTCTG 600
AAGCAGAGTGA AAGGGGCAAT TTTGTGATGAC GACAGATCA TAACTACTCT GAGAACTCTG 660
AAGAGAGAGG CTTGAGAGGT CACAGGAAA GTTTAGGAGA CGGAGATTTT CATGAGAGG 720
GTGAGAGAGG TGTGCGAGCA GTACTCTCCG CTCTCCAGCG CTTGAGAGG CATCTACTTC 780
ACCAATGAGT CCTCTAGCA GATAGACTTC TTGTACAGAT ACTCCCTCTA GTTTTCTCTG 840
GACATTTATC ACAGCTGCT ATACGAGAAC CCGAAGCTGA AGGTGTGAC CACACAGCA 900
CAGCGCTCTT CATTATTAAC AAGGAGCTTC TTGAGGTGG CTTTAAAGG AGTTGAGTGA 960
GGCATCTCTC ATCAGAGCA CATTACTCTT GCAATCTGTC TGGAGAGAT CAAACTGAG 1020
GGCAGCGTGG GGGAGGCCAC CTAGATTCGA GATTCGAGC ACTTCTGTGAG AGGAATGAG 1080
ATTGTCTCTA GTGCTGCTTC CACGCCGAG ATCCAGGGCC TGACTGTGGA GAGAGCGGAG 1140
GCGGTGTGTA GCGTGAAGT CTTTCCCGGG TTAAAGGACT TGAATGAAA GGTTCAGGCA 1200
GACGAGCAAT TTGCAATCTG GCTGAGAGC AGCTCCCGGG AGCAGACTGT GCGCTACTTC 1260
TGGATGTGAG AAGACCTGTC AAGACCAATT GCGAGGGCA TGCAGCGCTT GCTGCTGATC 1320
CAGGCTTTTC GGGCGGATGG CTTGTTGGCC ATGGGCCACA TTTTGTGTTT AAGAAACCTT 1380
GGGAGTCTTT TCAATGTCAT CATGAGAGG CCGCTGAGC TGACCCAGAT TTGAGSACA 1440
GAGGTGAAGC CCAAGCTCC TGTCTTAAATG TGTCTGTGTC CTTGTTATGA TGGCAGTGA 1500
CATGTCCAGG ACTTGTGAG CAGAGAGAAC AGCGAGTGA CTTGATTTTC ATTCGGCTCT 1560
GCGAAGGCTT TTAAAGAGC AGATAAGCA ATAAAGAGC CTTGTAAGTC GGGCAGTGG 1620
GTGATGCTGA AGATGTGCA TCTGCGCCA GGTGTGCTGA TGCAGCTGGA GAGAGATTTG 1680
CAATTCCTCC AGCGCGATCC CTGCTTCCGA CTTCTCTCTA CCAATGAGAT CAACCCGAG 1740
GTGCTGTGA ATGTGCTCCG TCGCGGCGCC ATCTTTTGTGT TCGAGGCCAC GCGAGGATG 1800
AAGGCCAACA TGTGTAGGAC GTTTCAGGAG ATTCGGCTCT CAGCGATATG CAGTCTCTCC 1860
AAGGAGCTGG CCGCTTGTGA CTTCTCTGTC GCGTGTGTTTC ATGCGATCAT CAGAAAGGAC 1920
TTAGGATAGC CACCACTGGG GTGTGCAAG AAGTATGAT TTGAGAGATC TGACCTCCGG 1980

2040 TCAATTTTGG ATAGGTGGA CAGTGTGCTG GATGACAGCG CCAAGGGCAG GCAAGACATC
2100 TACCGGATA AGATCCGCTG GTCTGCTACTA AAGACTTAA TGGCCGATTC CATTTATGCG
2160 GGGCGGCTGG AAGACAGTTT TGAAGAGCTT CTGCTGACA CTTTCTGGA GCGCTGTTTC
2220 AAGACAGGA GTTTGAGCAG TGAATTTTAA CTGCGATGCA AGGTGAGCG AGATTAAGAC
2280 AATTCAAATG CAGATGCGAT GCAAGCGAGA GCAATTTTGT CAGTGGCTGG AGTTGCTGCC
2340 CAGACCGAG AGCGCTCTCT GCGTGGGCTT GCGCAAGAC GCGGAGAGAG TCTCTCTTAC
2400 CAGCAGGCT GTGAGATGA TCAATTAAT GCTGAGATG CAGATTTTGG AGATTAAGGA
2460 CAGCTGCGC TACGAGAGA CTGAGAGAA GACAGAGACA GATTCAGCT CCAAGCGGCG
2520 CCGTCTCTGG ATGCGAGAC TGCAGACAC GCGCTGACAC TGGCTGACC TCAATCCGCA
2580 GAGCTGAGC CAGCTGAGC GCAAGCTGGA GATTTATGAG GATCTTTTGT TCAAGTCTTT
2640 TGAAGAGAA GTGAAGATGG GCGCAAGCT GCTTCAAGAC GTTGGCGAG ACCTTGCAGA
2700 TGTCTCTCAG GTGTGCGAG GAAAGAGAA GCAAGACAC TACTTGGGCA CCGTGAATCA
2760 CAGCTATGG AAGGAGATCT TGGTGTGGAG CTGTCTCCAC TACAGGCTTC CTGCGGAT
2820 GAGCTGCTC CAGTGGGCT CCGACTTCTG CAGAGAGATC AAGAGCTTC AGACATCTC
2880 ACTGCGAGCT GCACTGTGTT GCGCAAGGA GCTAAGAGAC ATTCAGCTGT GCGTGGCTGG
2940 CCGTGTCTGG CCGTGAAGCTT AGATCTGCTC CACAGCGGAG TATGTGCGCC AGGCGACAG
3000 CTGTGCTGG GAGGAGCTCT GCGTGAAGT CAGCTGACC AGCTGAGAG GCGGCAAGCT
3060 TGAAGCTTC AGCTTGGAG TCAAGGCTTT GAACTTCAA GCGGCTTCCC CTGAGCGAG TCGCTGGCT
3120 CAGCTGCTA CTGTGATG CCAATCTCAC GCGGCTTCCC CTGAGCGAG TCGCTGGCT
3180 CAGCGAGCA AAGAGCGGA AGAGCGCAG TGTGTGTAAC TTACCTGTCT ACCTGAAGTT
3240 CAGCGTGA GACTCATCT TCAAGCTGGA CTTGAAATTT GCTAAGAGG AGGATCTCTG
3300 CAGCTTAC GAGCGGCTG TCGAGCTCT GTGAGAGAG TAACTTTTC TACGTGCGCC
3360 TTTCTGTAA AGTGAAGTT GATTTTAACT AATTATTAAT TTTTAAATA TTTGGAAGT
3420 CTGAGCTCT GAAAGAGAG TGTGTGTCT GAGTGTGAG GAGCTGAT GAACTCTGAC
3480 GGTGTGAGT GGTGAAAT GGAAGATAC CAGAGATAT TTGCGAGGC CAATGCGTG
3540 GCTCTTGA GGAATAAA CACTAGCAT GAAAGAGAA AAAAACTTA CAAGCGCAA
GG

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 883 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

10 AGGTGATTTT ATGATAGCTT GTGATGATA GAGGGATTA TCTGTTTACA TTCTGTTCTT 60
CTGATATCAC TCAACAGGCG GTAACTAGCT GACAGATCTT ATTCAAAAGA 120
GCTCTTACAG CAACCAACAG TCTCATCTTC CCAATGATTA GATGACGGCG CCTTGAGATA 180
15 AGCTACAGGC AACACACTTT CCGCTTTTCT CTTGGCCCTT GATTCCAGAT GCGGATGAA 240
GCGACCGGAC GTTATGATGTC TGACATCCCG GTGCTGAGGA CTAAACGCGCG ACCCGGAGAT 300
CGTAGATTCT GGTGTCAGCG ACTGAGGAG GATATACAGT CCGTTATCCG GTATGTGAG 360
AACAACAGA ATGCTGACAA CAAATGGTTC GAACTGAGT CCAACAGAGA AAGAACTCGG 420
TGTTTGTGAA AATCTGATA TATCATGAC CTCCTGAAAT ATGAGTTTGA CATGAGTTT 480
25 GACATCTTA TCAATATCC TACTACTGCC CAGAAATTTG CAGTTCTTGA CTGAGATGA 540
AAGACAGCA AGATGATCAG GGTGGGAAA NTAATCTTGA CAGATCATTT CAACCTTTG 600
TGGGCGCAGG AATATGCCA AATTGGAAT AGCTCATCTC ATGCTCTTGG GCTCGGTGC 660
30 ATGCTGGCA GTGGAATCC CTATGCTGAT TGAGAGGGCG GTGATTCGAC ACAAGAGAAA 720
ATGCAACAAA TGAAGATCA AGCCACTGAG CGAGCGGACA GGAACCTTTG ATAGGCTAGG 780
35 ATACTATTTT CCTGTGCATC AACTATTAAT CATCTACTGT TTCCCGCGAC AACCTCCCAT 840
CTACTTTTGA CTAAATGATG CATTAGCAAT TTGCGAGAGA ACA 883

(2) INFORMATION FOR SEQ ID NO: 65:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1541 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

50 GCGACGAGCT GGCCTTACC CTGGCTCAT CTGGCTTACC AGGGAATCTA AACGCTTCCA 60
GATTCCTCGG AACATGCCA CCGGCATGAG CCTTCACAAA GAGAGAGAAA ATACATTTT 120
55 TAAAGGCTGG GCTGTAGAGA CAGGAGATA CAGGAGAGGG GTGGATGACC CTGACCCGAG 180
TAATATGAGG GCGGAGCTGC GCTGTGCTCT CATATAGAGC ACGAGTTTCA ACCTGATATA 240
60 TGATGGGACC AAGAGAGTTC CATTAGACC AATGACATA TATCAAGTGT GTGACATCCC 300

5 TCAGCCCGAG GGTCTGATCA TTAAACCGAG ATCCAGGGG TTCTGCTCCCT GCGATGAGA 360
GGAATATGAT GTGATGAGAG AAGTAGAGAG AGATGAGCTG GATCATGTCC AGCACTAT 420
TCGATTCAG GACAGCTTCC CTTTCTGAAA CATCAATGCT TTCTCCATGG CCGCAGCCAG 480
TGTGGCCAT TGCATGTGCG GAACTGCGAG CCGGAGGCA GTGTGGGCGA AACTGAAAC 540
10 CCTGAGATG GAAGTAACCC AGGCACTTAT AAGCCCTTC TATAGCTCTC CAGAACTGTG 600
GATCAGCTCT CTGCCATGTA CTGACCTGGA CATCAAGTTT CATTAACCTG GGAAGAGATA 660
CGGCGAGACC ATGACCTGTA GAAACCTTCA GGGTGGCGA CTCTCTTATG GGAACCTGGG 720
TCCATGCTT GACGAGGAG AGCTCTTTGG TCCGTGAGN CTGAGCGAG TCMAATTTCC 780
AGTCTCTGAG CATATTAACA ATGAGAGGCA GAAGCTGTTC ACTAGCAAGC TCGTGCACGT 840
20 CATGAGCAGA GAACTATCC TGGAGGTGAG CCGTCAATCC ATTTATGCCA TGAAGGCTGTG 900
CGATGCGAG GTTACTACTGT CTGGGCATG TCCGCCATCA CTGTGTTGCTC CCAACCTCAT 960
TGAGAGACA AAGAAAGTCA AGCTATTTTG TCTGGAACA TTGCTTACCG ATCTCATTTCC 1020
CGACGAGAAA GACAGATGAG AAGAGCAGCC ACCGTTTGAAG ATCTACTTAT GCTTTTGGGA 1080
AGATGCGCCA GATGGGAAC CATGGGAAG GAACCTCATC TTGGTTGAGG TCATTCGAGT 1140
30 AGTGGCTGGG ATGATCTAGG AGATGTTTTC TGGTGATTTT ACAGATCTCT TTGATATGCG 1200
CAATGTCGCG CTGAGATCT CAACCGAGA CATCAAGAT AACTGCTTG CTACGCTGAA 1260
CGAGCTGAC GGAATCTTTC AACCAGAGA GAUCTGGCAG CCAATGAGC CGACCCCGAG 1320
35 CATGCACTG CCGCTGCCC TCGCTGCCA GTATTTGTGA ATGCAATCTT CTTCCTTCTC 1380
TTTTTTTAAA TATGTATCAT ATGCAATTTT TTATTTTATA GATTTAACA GCTTTTAAAT 1440
40 CTCTGTGTTT TGTGACATG TTAGAGATTT GTGATCTCC AATATGCTT AGATTTAAG 1500
CTGATTTAAT TTATGAAAA AAAAAAAAAA AAAAAAAAAA A 1541

(2) INFORMATION FOR SEQ ID NO: 66:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

AGAAATGAAA TATTAGAGGG TCGTCTCGCA GCGGAGAGAG AGTGTTTTCT CCGCTGAG 60
AAGGCTCTGC TCAAGCCCTCA GAGTCCCTTC CTGCCCCAGC GATACTGGCA CTTTAAAG 120

GAAGCTGACC GCAAGCTGTC CAGACGATTT GCGCCCGAGA AGATGGGGAG TTCTGTCTCG 180
CCCTTCTGTG TCTGCTGAC CTCACCCAGC CTAGAGGGA GGTGCATTTCA GGTAGATATT 240
5 GCTCTCTATT CAAATTTCTG GGGCTTTGGG CGGAAACAG CAGACTTTGG GCTGTGTGG 300
GAGACTCTTC CAGACAGGA ACCCCAGAG GAGACAGAG CTCGCATATC CTCCACAGCC 360
AGGCTCTGG CAGAGGTGAT TGGACTGAGA ATTTGGCCAC AACCAATTTG ATCTGTGCTG 420
10 GACACAGAG CCGAAGAGCC TGGCTTTGTC CCCATGTGGG AGCCCTGTCC TCAGCCCTCT 480
TGTCCCTGTG AGCTCATGTA ATTTCCAGCA GGTGCCAGCA GCTCTGTGAC TTCAAAATTT 540
15 ATATATTGAG AGAGTTGAG AGTATATCAG AGATATTTTT GGAAGGAGT TGGCTATGTC 600
ATGTCTAGTT TCGAATCTTC TTGAAGTTT ATTTTATTTA TTAGAGATT TAAAGAAAT 660
AAAGCTTAC AATATCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 720
20 AAAAAAAAAA AA 720

(2) INFORMATION FOR SEQ ID NO: 67:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 629 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTAAGGAATT CCGCCGATC CTCGCAAGTA ACATGACTTA AAGAGACCG GAGATCTCG 60
GCTTCTCTCT AGAGATCGAT GGGCTAGAGG AGAGCTGTC CCAAGTGTGG AGAGACTCG 120
40 AGCCCTGTGA CTCGAGACT CACAGCCGGG AGCTGAGCCC AGAGCCGAGG AGGTCCCTCG 180
AGAGGAGAA AAGACGCTTA ATGACAAAG CTTCCACTTA CGAGAGGAA CTGAAGTTTC 240
TTCCGCAAGA GAACCGGAG AACATCTGC TCTGTGTGGC CATCTTTATC CTCCTGAGCC 300
45 TCTCTATGTC CTACTGAGCC ATGTGAGCT GGCATTTCCC CACACACAG ACAGGCTTCC 360
ACTTGGCCCC TTGCTAGGA TCAAGCAGGC ACTTCAGCC TCATAGGAC CAGGTGCTG 420
50 GGGTGTTCCT CTCGCAACT AGTGTGAG CATTGCTTCC TGGCGGCCCA GGGCTTGCCT 480
CCCTGGCCCT CTGGGGGTT CCGGCTTCC AGAGGAGAT GGTGCTGCTC CTCTGCTTAG 540
55 CCGAAGGGAG AGGCATPAA GAACCAAG CTGAATAAAA AAAAAAAAAA AACTGTAGG 600
GGGGGCGGCT ACCCATGCG CCTTTCCTG 629

60

(2) INFORMATION FOR SEQ ID NO: 68:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1751 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CTCTTAGCCG GCGGGGCGAG GCTTCCAGGC GGTATAGGCG GCAAGCCAGG CCAAGGCGCT 60
GTTACCCCGG GGTGTGCGGC CTCAGTCTGC GGCATATGGG GGTGTGCGGC GGTCTGTGGG 120
15 AGGGTGTATC ATGGGGGCCC CGGGCTTGGG CAGGGGAGCC GTGTGTGCGC GATCACTAC 180
ACACTTCCAG CTGAGCGACC TCTCCAGCGG GCACTGCTTC GGGGAGACA TGTCTCGGGG 240
CAGAGAAATT GGGTGTAG CCAAGGCTTT CATTAACCA GGAAGACTCA TCCAGATCA 300
20 TGTCACTACT GGGTGTGCGC TTCAATGAGT GAAATATCT ACCCATCTTA GCTGTCTGTT 360
GATGTGTTTT CAGAGAGAC TTCCAGCGCC AGAGGCCCTA GATAGAGCTT ATCAGATCCA 420
25 CACATGATTT AACCTGATG TGGCTTTTGA GTCATTAA CAGCCCTTTA CTGTCTGCTG 480
GATTCATCCC GCAAGTGGCC GAGTCTATTA CATTAATTC AACCTTCCA AAATGTGCG 540
30 CATTAAGAC CTGACTGCGG AGCTCTCAT TCAAGGTGAG GATGATTAAC CAGAGAGCGT 600
TATCAGAGA CTAAAGCTT ATGAGACCA AICAAAGCCA GTCTGTGAAT ATTACAGAA 660
AAGGGGTTG CTGGAACAT TCTCGGAGC AGAAGCCAC AGATTTGCG CTAATGTATA 720
35 TGTCTTCTTA CAATCTAAG TTCCAGAGG AACCCAGAA GCTTCAGTTA CTCCATGAG 780
AGAAATGTGT GTAACTATTA ATAGTAGAT GGCAGAGCT CTAGTCTCTT GCATTTAGAA 840
40 GCTGTCTTTC CTAGAGCTTC TACTATGTAT GATTTCTTG AAATATATAT TACTTTTATT 900
TCTACTGATT TTATTTTGA TACTAGCAT GTCCAAATG ATTGAGTAC TAAGATCAT 960
GTTTGAAAT CATCTAGTGT GTTGTATGCA GTTATCTCA AAAGATCAG CAAATCTGTA 1020
45 ACCTTTAAA CATCTGTTAG AGCAAAATTA AAGAGCATT TGTATGTAT CTACTTTT 1080
GTTCTGTTAA TAACTGTGTT ATAAAGTTTC CATATTTTTC TGGAAAGTTT AAAAAAGTT 1140
50 ACATGTCAAT TGGAGAAAT ACATATCAG AAATTTGCG ATAGATGTAT GCGAAAGAG 1200
AGATTTCCAG CATGTGTGAA CATGTGAGA CACTATTTAA AATTCAGAA AGAAGGAGC 1260
TGGATTTACA GATTTATGT GAGACAGAA TTCACTGCTG CCTTACTCT AGGAATGTA 1320
55 TATGTTAAGC ATATATCTG TATTTATTTT GTCTTTAGC ATACTTTTCA TTTACTGAGA 1380
ATTTTCATTT TCTATTAAG ATGTATCAT TACATATAG AAAATATTA CTTTAGATG 1440
60 ACTGTCTTCC TTGAAATAA CTTGTACTT GAGGTTTATG ATTTGTGTCA AAAATTGACA 1500

TAAATGCTTT TACAGCAC AAGTTGAT GATTTTCA CAAATCTAA TTAAGTCTA 1560

TTTTTCACT TATGACTCAG GTTAAGAAAT GTTTTTAGG ATCTACTGCG TGGTTTTCT 1620

TTTTTATCCA ATTGTGTGAT CTGCTCTGAT AATATACAG TTAATGTACC ATCTCCCCG 1680

CGATTAAAA AAAAAAAAAA TCGAGGGGGG GCGCGGTACC CAATTTCTCG 1740

NAATAGGAG T 1751

(2) INFORMATION FOR SEQ ID NO: 69:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GGCAGCAGAT TATGTATTAA AATGTTTTG AATGTGAAA TATTAGATA TTGTACTAT 60

TTTACCCAC TCAAAATCT CATGCGAAA TACTGTGCA TACCACAGT ATTGTGAAA 120

ATATACAGT GCATATCAC AACTGTATCA GACTCTAGTA CAGTTTGGC AATCAGGCA 180

CAGCTAAAA TTCAAAACA AATCTGCAC ACAAACAGC CAAAGGTGG GGGTCAGAA 240

GCTCTGAGT GTACTAGCT GTAGATGCT ATGCACAGT GCGAGTGTA GTGTGCAAT 300

CCAGAAAAA CTCGAGGAG CCCAGCTTT CACTCTGTT TACGCTGAG CTCGTGTGAA 360

GCAGAGGTG AAGCTTAGG CAGATTTAG CTCGTAGGC ATTCCACAC ATACACACA 420

ATCTGCAAA GCATTAGGA AATCTGTGA CTCCTTAGTG TTCTGAGCC AGGAGAGCT 480

CCTACTGAC TCGACTTAAA AATAAAA 508

(2) INFORMATION FOR SEQ ID NO: 70:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 245 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TACATGAGC AAGAGAAAT TTCCAGATT TCTAAATTC TCGAAGAGA ATTTCCTCA 60

GATTCAGAT TTGCTGTGT CTCAGGTGA TGTAGAGGC TTTTTCGCC TTTTGTGCTT 120

TGTCTCACT CAGCTTCTT CTCCTAGAT GTCTGTGAG CATGATGTC TGTACTTAG 180

GCATTTCTTT CACTATACA AGGAAACTG CAGGTTTAA AAAAAAAAAA 240

NKG 245

5

(2) INFORMATION FOR SEQ ID NO: 71:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 361 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATGTTCTCA TGAAGTACA CTGTGCTTC TCCAGTATT GCTGAGCTT CATATGACT 60

CCGACAGCA CAGCAATAC AGTACTTAC CTGTGCTT GATCTGAC CAGCATGCT 120

GGAGAGGA CGACTGGGC ATGTACCTA AATCTGTTA CAGGAGAG ACTGCCAG 180

TCAGACAG TACGACTTC CTCAGAGGT GATCACTGT CTCCTGTAT CCACTATT 240

TTTTTTTTT TTTTGAATA GATTTTACC CTGTTGCC TGGTGGAGT GCATGCTG 300

GATCTGACT CACTGAGCC TCTCTGCTC CCGGTTCAA GTGATCTCC TGGCTCAGC 360

T 361

(2) INFORMATION FOR SEQ ID NO: 72:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 713 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

AGGATCCAC AATAGAGAC ACTGTAGTA CATTTGCTC TGTCCAGG ACCGACACA 60

TTGATCAAT TTTGTTGTC GTTATATAT TTTCTGTAA AATTTGTCA AAGTTTCTT 120

TTACTAGAT GATTTTAA TACTGGGAG TGTCTGAAA CTATTAAGT GTCACTACTT 180

AACAACATA CTTATGTTT TTTTGTGTT TTTTTCAC TCAATATAA TCAGGAGAG 240

TTAGCCAGC ATCTAGCAT TAGATCTTC TTTTATTTG TCTCTAAG ATATGCAAT 300

TCCCATACA GCACAAAC AGCAACAAA ACNTTCTTA ATATCATTT GATGACTGT 360

AAGACTGTC TTACTTGT GTCCAAATA TTTATGTGT ATATATATAT ATATATATAC 420

AACAACAC ACTATATAT TCACAAATA AGCAAAATA TACATGCTT TTCATTTT 480

GTCTTTCCT GTTAGCATT TATAGCAGA ACTATATGAC AAGTTTAGT GATCTTACA 540
TATCTTAAT TCAATTAAT GTAAACAGA TTAAACAAA CAAGAAACT GTCTATTGGA 600
5 GTAGATCAT GCTTCTTAT ATATTAATT GCTTCGGTT ATCTCTAAA TCCACATTA 660
TACTGTATC TCAATTTTAA TATTAAGAA TACTTACTT ATAAAAAAA AAA 713
10
(2) INFORMATION FOR SEQ ID NO: 73:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 73:
20 GAAGTCAGA GCTGTCCAT CCTCAGAC GTTTTAGT TCTCCAAAT TAGAACTG 60
GGACTATGT GTCTGGCA ATCAGGTC TGGAAATGG CTCTCAGGC TCTGTATGT 120
25 GAGACGTGG TCACTTACC AGACATCAT CTGATTTTAA GCTCAGGCT ATTCACAT 180
GCTGGCCAT GCTATGAT ACAAACAA AGCAAACTCT GCTTTTAG TTAGCAAC 240
30 CTGATAGA CAGTATTT CAGATCTT GGAATAGCA GTTCTGCAT TTTAAATGG 300
GACTCAGAA GTAGCTGT ATAGTTGCA AATACAAA ATGTATCTT TCATCAGAA 360
AGAGGCCGG TCCCTGGAC TTGAAAGT ACTGAGATC ATAACTTAA TGAAGAAATG 420
35 TAGGCTCTGT GATGTATAC TAAATATAG GTTAGGAG CATATGAC GCTTTGTCA 480
ACTAGCTTC ACTAGAGA CTTGTCTCA TAGAGATA TCTTTTAAA GTATCATTT 540
40 TCCACAGTC ATGATGAGA AAGTTTCAT TCACAGAA TCTGTATGT CACATAC 600
AGCTGACAT ATATACAT AAGTTTCT GTAAACAGA GTTCTTCTC TTCCATTTCA 660
45 GGATCAGTC AGACATAA TATTCATGT TTCATTTAG AACTGATTC ATTTTAGAA 720
GGAGATCTGG ATATTTTGG AGGTAGAAA TGAAGCAT TTCTGGCAT CTTCCTCAA 780
AAGTCATAT ATATACATA AGTATAAA AGGCTCTCT TCACCTCTT TGTTCGTAG 840
50 CATGGCTAC ATACTGCT CC 862

55 (2) INFORMATION FOR SEQ ID NO: 74:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4602 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

5 GCGAGGGGCG GTGGGAGCA GCGCGAAGC CGCGGCTCC GCTTCGGCG CTTAGACTA 60
GGGCTTGGG GACGGACAG CCGCATTCG GGGGAAATG GAAGAGCCA GACCCCGCA 120
GCACAGCAC CAGGAGGGG GAGAGGGGG CAGGCGGCT CCGAGGACC CGGACAGCG 180
GCCCCCTGAG GGGTTCGGG CGGCGCTCC AGGACCACT AGAGCGGG TCTGAGGGG 240
AAGTGGGAC CAGGCGGGG CGGCTGGC CCGCGCGAG CAGCTGTTC CGCGGAGA 300
15 AGGCGCATTA TCCCGCGCG CAGAGCAG CCGAGGCGC AGGCTTCGG AGTTCGGCG 360
GCACAGCCC AGGCGCGCA GTCCGCTCT CAGTTTCTG CGAAGAGAG TCCCGGACTC 420
20 CTATCTTAG AAAAGTTAC AACTGATA GATCCAGG AGAAAAA GGAAGAGAC 480
GATTCGCC TCCCTCAGA AGTTTCAT GTTCATCTA GACTAGCG GGGCTGGCT 540
AGTATAGA CATCTGTTT TGGCATCTT GATACAGA ACACCGAAG CTCTCGGTC 600
25 AGAGCGGTT CATTCAGT CATTTCTAG TCGAGCAA ATACAGCA ACTGATAT 660
GATTTGAG AAGACATCA GTCTCAGT GCAATTA TAAGAGAG AGAGAGAG 720
GAGAGAGA TTTTATCTG TGAAGAGAG ATACATCTA AGATATTC AAGATATAG 780
30 ACTGAAAC CCGACTTAG AAGGAAAC CCAAGGAC GAGAAAAATC AGGAGAGTA 840
AAGAGAGA AGGAGAGA GGAATTTAA GTGAGTAG AGTTGAGT GAAAGAGAG 900
35 GAGATGAAA TTAGAGAGA TGAAGACT CAGAGAGA GAGAGAGAG AGAAAAAT 960
GACAAAGTC CAGTTTACC CAAAGAGA AAAAGGCTC CATTCCATA TGTCCCTTGT 1020
40 GAGATGAG GATTTGAG TGTCTTCC CATCTGCT ATTTGCACA CCACATTTAA 1080
TACCGCAT TGTCTAGAA GAATTTATA TGTCCCATC CTTCTTGG AGACTCTTC 1140
45 AGGTTTCTG AGCACTTCT GGCATTCG AAACATATA CAGTCAAG GATTTATTC 1200
TGTGATATT GTCTCGGG CTTCAGAT TCCCATATC TGGCATGCA CCGATTAAT 1260
CAGCTGGG AGAGCATTA CATTTGAGA TCTGTGAT TACTTTTGA CAAGAGCAT 1320
CTCTTAATG GCATATAG AACATGAT CAGACTCTT CTACCATTT TCTTGCATA 1380
50 TCTGTGCA AAAATTTAG AAGAGACA GGTAGTGG ACAGAGGA AAAAGGACC 1440
CTGAGTCT CATTTGAGA GCTCTGCT CCAATGAG CGGCTTATC ACAGAGAG 1500
55 ATATCTTGG CACTAACCA GATCTCTGA CCGAGCTTC AGATGCTAG GCTCTCTC 1560
TTCTTCTGA GCGCTTGGG AACTCACT CTGAGATG CTTACTTCTTA GAGCTGAG 1620
60 GATTTTAAA GTCACTATC AGTGGAGG AACGGTGA CTTGATGCT GATGGAGA 1680

1740 TCTTTGTGG AGCGCGACG AGTGGAGCA CTGAGGECT GGTATGAC TCGATATAC 1740
1800 TCGGTCTAC CACAGAGTT CTGATTAAG ATTGAGATC TCGCGAGCT TATGGACAG 1800
1860 GAGAGTTGG GGCATGGAC AGCTGAGT TGTATTTAA AGTTAATAA GGCATATAA 1860
1920 AAAATCTAA GCTATTAATA TCTATGAAA TAACTGAGG GCTCTCTTT TCCATTTGG 1920
1980 ATCAGACAC ACATATACAT ACACCTCCA CTTCCCATC CCTCTTCTC CTTCTTTGC 1980
2040 TCCCTTATA AAATGATGT TGTCTTACC AGAAGGTAG ACATATAAG AGCGACGA 2040
2100 GCTCTTAAG TGAAGGTAT TCTATATCT GGTTCAGCC ATCAGAGAC TTCTCTCTA 2100
2160 TCGGCAATC CCGTTTCCA ACCTTAATCT CTGATGTCT CTGATACG TTTTAATTT 2160
2220 TAAATCTATA TTAATCTCT CTAAATGCT TCTCTCTCT TACTCTGCC CTATGTTCT 2220
2280 GGCCTCTACC CCGTGGCA CACTATCTT CAATATCCT AAATCTATA TCTCTGAAT 2280
2340 CATATCTCT CAGTGGCTT TAAAGAAC TGGTCTCTAG CACTATCAA ATCCTATCA 2340
2400 TACCTTAGT CTTTCTTGA ACCCTTTTA GCGAGNAT TTAGTTTCAI GGTAACTAT 2400
2460 ATGCTCTTG AGTTTATAC AGTTTGAAA CTTAAGAT TTGAGAGGT GAGAGGTT 2460
2520 GTTCAATC TAAATATAC ATGATATAT GTTCTTTTG AATTCTTTC TTTTCTTTT 2520
2580 TTTTCTGCC TACATTTCC TTACATTTCC CTGCGGCC ATCTCTGCT CTTCTCTTT 2580
2640 TGTCTCTCC TTGCTTTAT CAGTCTATC CAGCTCTCT TTAGTGAAG ACACCTCT 2640
2700 TATGAGGA ACATATCTA TGAATCTTA AATTATAT CTAAAGAAC TGGTCTTTT 2700
2760 CCGCTTTAT ACATCTCT AGAGGAAA ACTTCAATG CCAATGTTA TAATCTTAT 2760
2820 TAAATATCC TTTGCTTTC ACCTAATTT CTGGCATCA CATTTCTT GCGATGAG 2820
2880 TTCTTTTGG GAATGATG CTATATCTG TTAATGGAG AGAAGGTTA GTTCTTTCT 2880
2940 ACAGGTCAT ACCCGAGAA GCGCAATC CTATTTGCC TCAATCTCAG GTAAAGATTA 2940
3000 ATTCTATCC TGTGCTCT AGAGCTAGA ATCGAAGCT TACCTATTC ATTCTTAT 3000
3060 GTCTGAATG CATGATGCT CTGGAAGA ATGAGTTTT CTCTGAAAA AAAAAAATA 3060
3120 CAGTTTCT TTCAATACA TGGCTATCA ATTCTTCAA AGATCTTT TTTCCAAAA 3120
3180 AGAGGTTAA CAATATCTA TTTTGAAG AGGCTTACT TATACAACT ATGTCTAGCA 3180
3240 TTTGGATCC CAGGAAACAG AGATGAGAC ACCTCAATC ACCAGTCTA AATGCGTAT 3240
3300 TTTTCTTTT CAGATCTG CAGATTTCC ATTCTCTAT AATATGGGA TAGAATATG 3300
3360 AATTAAGTA GAGGATGT AGAATATCT TTCTGCCAA CATGTTTGG AGTCACTTT 3360
3420 GGTATATTA CTAGTTTGA AATATGAGA TTGTTAGT GAATCTACA AAATGTTCT 3420
3480 CTCTCTCAG GTCCCTTTTA CACTTTTGA CTAACTAGA TCTAATATC ACATTAATCT 3480

3540 TTTTCTAC ACTATCTT TGTCTCGTA AATTCTATT GCACTGTTA GTCATCAAT 3540
3600 AATTAGCA CTAACAAA AGCAACTGC ATTATTAAA ATCTTCTCA CATCGGAA 3600
3660 AATGATCT CTTTCTCAT ACCTCTCTC CACAAAAA ACACTAAT GTTCTCTA 3660
3720 AATGAATCT TGTCTCAT TTTCTTTCT TTTAGGGTTC AGGACCTC TTTATAGTA 3720
3780 CAATTTCT AATTAATTT ATTGACGAG TTGCAATAC TAAATATTT TTTATAGAT 3780
3840 TTAATTTTT CTTTGTGTA AAGGATGCT GCATATAGA GTTGTGTAA TTAATATAT 3840
3900 TGAAGGTTT CCGTCTTTC CTTCTCTC CATATGCTT ATTCTCTTC CAGGAGCTC 3900
3960 TTTTAATCT AAGTTCTAC ATTCTATCT CTATGCTAA TTCTGTACC TTTTATATA 3960
4020 CTTCTCAC TCAATATTC CATCTGAAT TGTGTGTTT AAATCTCTAA ACTGTATTC 4020
4080 AATAGACT AATTAAATTT TCTGGAGAT GTGCATGCT CTTCTTTTG GTTCCGAG 4080
4140 GTTGTTTTC GTACTAGA CTTCTGATA TCTTCAAG AATTAGGA AACATGCC 4140
4200 ATGCGCTGG CACTACTGG AGTAATATA AATATGAG GTATAGACTA GCATCCAT 4200
4260 AGAGCACTG AACTCTCTT GTACTCTTT GCGAAGAG TATATGAT GTACTACAA 4260
4320 TCACTAAG AATATATAG TCTGTGTTT TCAATACCA TTTTCTCTC CTTTGTCTT 4320
4380 TTTCCACT TCAATATAC TCAAGAAAT TCAAAATG TAAATGATCA AATTAAATA 4380
4440 TTTTATCT TAAAGGCTT TTTGCTCT TGTATGCTC AGGACCTTC TCTTTCTG 4440
4500 GAGAGAG GTATGTTCT GAATATAGT TCAAAAGTT ATGTAAAG AAATATAT 4500
4560 AAAAACTA CTTTCTTTT CAATCTTTG TTTCTCTTA TTCTAGTAA GCAATATTA 4560
4602 AATTAATAT GTAAAGAA AATTAAGAG TTTCTCTAT GG 4602

(2) INFORMATION FOR SEQ ID NO: 75:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1255 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1x1) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

60 GCGCGCGCGT TCTCTTACA AATTAACAGA ACCGCACTG CCAAGGAGAG 60
120 GCTTCCACT TTCATATGAG TCCCTTGGG GAGCTAGCC TCACTGAT CATCTGCA 120
180 AGCGCACTT TTAATTTTA TTTATTTTT ATTTTTTTT TAGATCTTT TTGGGCTTC 180
240 ACTTCAGAG CAGTTTITA AGGACACCA GAGCGGAGC CTGCTCTGAT TCTATGCTT 240
300 GCTTCTACT ATAGATATA TTGCTTACT TCAATTTCA TCTTTTAA CAACTGTGT 300

5 GCGAAGAGT ATTGACCTT TTCGAAATTT CAGATTTCTGC CTCTGCGGAT AAATATTTTC 360
CAGCAATGAG TAACTCTCTCT CACACTCTG AAGCTCCAGA CAGAGGTTTT TGCACATTC 420
TTAGCACTGA ACTCTCTCTCT GATCTAGAT GATCTCTTCC CCTCTCGAT GACATCTCTC 480
TGATGATCA GCTCTCCAGC AGCTACTTTT GAAGGAGACA ATCAGATCCA AAGCTCTCTG 540
GCTGTTTATTT TAAATATCTA GTGTCACTTT CTGAGTACCC GCGCTTTTAC AGCTGAGTC 600
CAGGCTCTG TCTTTCTTAG ACCAGCTGCT TTGCTCAGAG CCACATTTCC ATTTCATCA 660
TTACTGCTTT CAGCTGATA GTCACTCTTTT TGATGCTGGG GACCAAAAT GTGATGATA 720
TATGACTTT ATGTATAGCC AGATCTATC CCGACCTTA GTCTTCGAAA TGTATATTT 780
TGATATATCT AGAATATGCA TTCTATCAT TACAGATTC AAATATTTCA AAAGATGTC 840
TGTCTTTCTC CCGAGCTCC CCTGTTGCC TTCAATTGAAA ACCACAGCG TGCATCTCT 900
TGTATATGCA GCGCTATGCA CTTGAGGCA GTGTGTATG CACTTCCCGC TTGTGTTTAC 960
ACAAGCTGT GCGTGTATG CATGCTCTCT TTTTTCACCT AATATATCAG CTTGAGAGA 1020
TTTGTATATC ACHTATATA TCCACTTCC TCTTTTATGAT GCGCATATA TAACTACTGC 1080
ATAATATGCA TACGCTTAT TTGATTTAAC TACTTCCCTA ATGATGACT TTTTAAATGTT 1140
TTCTTTTTTT TTCTTTTTTT GCTACTGCA ACGATGCTAT AATTAATGTC CTTATCAAAA 1200
AAAAAAAAA AAAAAAAAAA AAAAAATCCC AGCGGGGGGG CCCCCGGGAC NCAT 1255

35

(2) INFORMATION FOR SEQ ID NO: 76:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 475 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

5 GCGACGAG AAATGTTGA TTCTCTTCC TATTTAAGG GATCTCTCT CTCTGTATG 60
TTGAAACTT ACTTATGTA AGATGTTTTT CACATCTGTT TTCTCTTTTA CTTGATAT 120
CAGAGTATG CATCTATCA AAGTATGAT CTGTGGGATA GTTTTAAATCA GTTCACAAAC 180
CAAACTAG ATGTAAAGG AATGATGAAA ACTGAGACC TCGAAGAGG ATTTCTTTA 240
GTGACTTTTC AAAGAAGG AAGGAAGCTT TTTATACAC AAGAGATTT CTTTATTAAT 300
ATGAGGCTTG AATTCGCC TTGATATCA AGGTATATGC CTTCTTTCTT TTGATGAT 360
CTCTTTTGA CTCTAGGTG GAATATTTT GAAGTGTTTT AATATATATA GTTCTGTGA 420

60

5 AACCTACAA GATTATCTT TCTTAGAAT ACTTAACTT CTTACCAAT TAAAA 475
(2) INFORMATION FOR SEQ ID NO: 77:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 77:
15 TTCTCTCTGC TCTTCTACTG CACCGCACTC GCGGTGACG CTGACTCCC CTAGTCACT 60
CAGCGTCTCT GGCATGCGGT GCGCGCGGCG CAAACCTGCG TCGCGGCTCG CCGCGTGTG 120
20 GCTCTGCGCT TGTCTGCGCT GCGCTGTGCG GTCCCGCGCG CCGCGCGCGCG GCTCTCTGAG 180
TGTCTCTCGG CCGTGTGAAA CATGCAATAC GTGACCGCC AGACCAACT GACGCTGTG 240
25 ACGCTCTCGG AGATGCGCG CTTCGCGGAC AGCTGCGCCA AGAGGCGCGC GCAATGCGCT 300
GTGCGCGCTC CATTGCGGCC CCGCGGAGAT CTCAAGGCT KCGCGCGCGA CAGCGCTTC 360
TTCTGTGCGG AGCGCGCGCG CCGAGCGGCG GCGCGCTGCG TCGCGCTGCT GGTCTGTGCG 420
30 GCTGCACTT TCAAGGAAA AGTGTGTGTC GCGCGCGCGA AAAAA 465

(2) INFORMATION FOR SEQ ID NO: 78:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1907 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

45 ACHTCAGCC CACTACAGA TTCTTATGGA ATTCTGAG GTTTCAGAA GAATTAAGG 60
AGACCACTG GACAGATCC AGAGAGGCT AAGTGTTTTG GAAGAGATA TTAGAGAT 120
50 GAGAGATG AGTGGTAT: ACTCTCTCTCT CAGTGAAGAT AGCAAGTGC CTCATTTGA 180
AGCTCTCTCT CCAACACCA CTAGTATAT TGAATCCAA GAATACAGCC AACCTCAGG 240
TTTCTGTGCG AGTTCTGGA CAAAGAAC GCTTGTGAT AATAGCACT TACGCTAAG 300
55 ACGAAAGCA CTTACTGTC ATTTCAGCA CTTGAGCAG TTTTACTTTT CTACAGAT 360
GTCTGTATC TCGATGACA GTGCACTGC AAGCGATTC GATGATTTTC AGGAATGCTT 420
60 GTCCAGTTT ACTCGATATA ATTCAATAG ACCTTTAGCC ACNTGTGAT ATGCTATGCA 480

540
600
660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
1380
1440
1500
1560
1620
1680
1740
1800
1860
1907

5
10
15
20
25
30
35
40
45
50

TCTCTATAT GATTGAGTA TACTCTCTAG TATTGAAAT GACCGGATTT GTGACTATTT
TCCGATTTCT GCACTTACAA AGAAGTATA AGTCTATGAA TATGACACTG TCATCCGAGA
TCCAGTGGAT ATTCTATACC CTGAGATGCA ATGACCTCTC AATTGAAA TCACTCTAT
CAGTTGGAGT AGTTACATA AGACTCTTT AGCTAGCACT GATTATGAG GCATCTATAT
TTTATGGCAT GAAATGAG GACGAGATC AAGGTCATAT CAGGAGCATG AGAAGAGTGT
TTGCACTCTT GACTTTAAT TGAATGATC TAACTCTTGT GCTTCAGATT CTGATGATC
AAGAGTGAAG CTGTGCTCTA CAAATCTAGA CAATCAATG GCAAGCATGT AGGAAAGGC
TAAATGATGC TGTATTAAAT TCAGCCCTCT TTCCAGATAC CATTTGGCTT TCGGCTGTGC
AGATCACTGT GTCACTACT ATGATCTCTG TAACTATGAA CAGCCAAATCA TGTATATGAA
AGGACACCTT AAGCACTCT CTATGCGAAA GTTTGTGACT GTTGAGGAAA TTCTCTCTGC
CTCAAGCAGC ATCTAGCTTA ACTGTGAAA TGTAGGGAAG CCAATCTGCC TACGTTTCCT
CAGGCTCAT ATCAATGAAA AATCTTTTGT AGGCTGTGCT TCCATGCG AGTATATAGC
TTGTGAAAGT GAAATTAAT CTCTCACT GTACTATGAA GCACTTTCTA AGACTTTCT
AATCTTAAAG TTGATAGAG TCMAAGTGT TCTGAGAAA GACCAAGAG AGATGATAC
AATGATATTT GTATGCTGTG TGTGCTGAG GCACTTACCA GATGGGAGT CCAATCTCT
GATTTGCTCT AATGATGAG GTAAATTA GCTGCTAGAA TTGATATGAA GGTATATCTC
AATGATATTT GTACTGATC CTCTGAAAT AGATCTGAG CTGACATGCA GAGAGAGAC
AGAAATATCT ATGTGATCT TCTGCCAAA GTCAATCATGG GTTTTGGATT TGTATTGAT
ATTTTCTCT TTTTCTTTT TCCCTCTCTT ATGACTTTTG GCAATTTGG ATATCCGAGC
CAATCTCCA CCAATCATGT AATCTCATGG ACATGCTCTC TCTTGGTGT GTATCTATAT
TTTGTGATA GCGAAGAAA TCTTTTGA TAAATATGAA TAAAGAAACA ATAAAGTTT
ATTGAGCCAC AGTTGAGCTT GAAAGTTT TGTCAATGTC NCGAAGGAT AACTCTTTT
ATGAGATGAC ATATGAGAC TATATGTA CAGTGAATTA TTTGTAAAT TGTATTTTC
CAACTCTTT GGAATTAACA CATATCAATA TAAAGAAAT ATTAAGT

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 79:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

5
10
15
20
25
30
35
40
45
50
55
60

GTGGGGTGT CCCKCSGCC ACATCTCTCA TCGCTTACTT CATGAAAGCAC ACTCGGATCA
CCCATGACTG ATGCTTATTA ATTTGTGAAA GCGAAGAGAC CATTATCTCT CCGAAGCTTT
AATCTCATGG GCGATTTGCT AGATTTGAG GAGAGCTTAA ACAGGGTGT GACACCGAGA
ATCTTACAC CAAAGCTGAT GCGCTGGAG AGGTTTGTGT GACATGCTCT TCGATGAAA
GAAATGCTGC TCTCAATAG GAGACATGCA GGAAGGAGCA TCGATCTCTG TTTTCTTTCT
TTCTTTTCTT TTTTGTAGTT GCGATTAAGT TTGTGATGG AATCAACTTT GTTTAATCAC
TTTATTTTAA ACAGTGTAA GAGACTATA ACTTTGATG CAATTTGAT TCACTCTCCA
CAAACTGACA AATTAAAGAG GTTAAGAGAG TAATTTTCTT AGCCCAACAA TAAATATATA
ATACACTTGT TTTCTCTCCC TTTTCTTTT AGCTATTTG TAAATTTTAT GACTAAATAG
TCTGTGAGG TTCAATAGCC GAGATACTA CACTATTAA ACCAATTAAA AAGAACAAA
AGTAAATAGA AAGACATG ATTCAGCAG GCTCTGGATC AACTGTGGCT GTCCACACAG
AAGAACAAA CCAACCAAAA CCAAGCCCTG TTGTCTCAC TGTGTCCAG AGAATATCAG
GCACTTTAA GTGCTTAGAG ATCTCTCAG CCAATCTTAA AGGAGAAAA GGAATCTTT
GATTTTGTGT GTTTTCACT CTGCAATTTT TTTTCTTCTC TTCTCTCTGG TTTAAGAGAT
TTTTTTGA ATATGAGGA ACTGACCAAT ATATGCTTTC ACTGCTTCT TGTGATTA
TATGATTTT TAACTGTCA AATGATTAG AGCTGGCAGC TGAATGATAG ACAAATATG
CAATTTGCC ACTCTGAGA TAAAGGAAA TTAACATTA TATCAATATC TTTCTTTCC
ACCTTTTCTC TTTTCTTTT TTTTCTCTGA TTTGATCTG GTTACATGTC CATTAACCTT
GTACATATG TATATGAAA TGTAAAGAAA AATTAATTA TTAATTAAT TTTTCCGAAA
AAGAAAAA AAAAACTGA GCGCGGCC

(2) INFORMATION FOR SEQ ID NO: 80:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

55
60

AGAAATCAC ATCTTAACAA AAGATCTGT CTAAAGCAT ACATCTCTG TTGAATTC
ATCTTCCAC AGGACTTCT GTTTTAAGG ATGAGACTAT TCTCTCTTC ATCAGGAAA
GAGAAATGTT CAGGTTTATA GCGATGCCAC ACTTATATG TCTGCTCTGC TGAAGGTTTC

CTCCAGGACA GTTTCGTGAC AGCTCCGATT CTATGTCAT GGTCTAATCC TTAGGATATCT 240
CTTCTTTTCC TTTCCTGTCT CAGGATACAG CTGAGATATC ATTGCATATCT CATGCTCTTA 300
GCCCTTTACT GTGATTGTGT GGTTCGACTT TCAATTGTCT TACTCTTAGA ATCACTGTGT 360
GACTCTCAG ACTTCACTTA ACTTTGGAAA CTCTCTTTTG GAGGCTTTCT ATTTCCCTCT 420
AATCTCTCC TCCCTGACC CTAGAAATTT CCACCAAGC AATTATTTCA GTTATGCTCT 480
AGTTCTCTC ATCTAGTTGA TATTTAACA ATATCTAGTT GATATTTCTC ATTCAATTTG 540
ATCCAGAAAC CAGTATCTCT TAAAGAACAC CTCTCATACC TTCTGACCT AATTTTGTCT 600
GCTCTGTCT GTCCGCGCAT ATGTATATAG ACAGCGCAT CTTTTTTACT TTTGTAAAG 660
CTTATGCTCT TTTCGTATCT ATATCTGTGA AGTTTATAT GATCTGCCAT AATCTCTGG 720
GGACCTTTCT CTCTCTGTGA AATGTTACTA GAGAAACAC CTATATTTAT AGTCATCTA 780
GTTCGTTTTA TTCCACATGA AGGAATTTTC CAGTATACA CACTAACAA CTCTCCCTTG 840
ACTAGGGGGA CAAGAGAAAG CAAACTGAC CATTAACAA AATTACCTCG TGAAGATTTG 900
CATTAACAGA ATTAGTATG ATATCTGAGA CAGCATCTT AATCAATTTT GTTCTCTCC 960
TTCCAAAAA CATTTACTGA CTTCCTCTTG AGTAATCCA AGTTTCTTTT TTTATTTAAA 1020
AACTTGCTCT TCAATATG TTTCAAATG CTCTGTTGG CCATTAATTT GAATGATGG 1080
AACTGACTGA TAAAGCTGA CAATATGA CTCTGCTTAA CAGCAACAC AGTAATTTTG 1140
ACATGCTTAA TTCAAAATG CTATTTCTT TATTAATTTT TTTCCTTAAA TACATTTGA 1200
AACTATTTT CTGTATTTCA AGAGCTGAGA TCTTAGATT TATGTAGTAT TAAGTGAAA 1260
AATACGAAA TAAATTAACAT TGAG 1285

40

(2) INFORMATION FOR SEQ ID NO: 81:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1290 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

45

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TTCTCAAGCC CAATTTCTAC GGCACCGGA AGACGGAGGT CCTCTTTCT TGCCTAACCC 60
AGCCATGCTT CTGTCTCCA AGAAGCTCT GAGCGGGTG CCAGCTCCAA AGCATGCTAT 120
GCTGGATAAA TTGACCGGTG TTTTCTCTC TCTTCATCC ACCGTTCCCC ACAATTTGAG 180
AGAGTGTCTC CCCCTATCA TTTTCTCAG GACAGACTT AAGTATGCC TGACAGGAGA 240
TGAGTAAAG AAGATTGCA TCCAGCGTT CATTAATAATC GATGCCAAG TCCGAACTGA 300

60

TAAATCTAC CCTCTGAT TCAATGATCT CATACGATT GACAGAGCG GAGAGATTT 360
CCTCTGATC TATGACACA AGGTTGCTT TGTCTATAT GTTATTTAC CTGAGAGCG 420
CAATGACAG TTGTGCAAG TGAAGAGAT CTTTGTGGC ACAAAAGGAA TCCCTCATCT 480
GTTACTCAT GATGCCGGA CATTCGCTA CCCGANTCC CTCTCAAGG TCAATGATAC 540
CATTCGATT GATTTAGAGA CTGCAAGAT TACTGATTT ATCAAGTTTC ATTCAACCG 600
CGAGGTGTC TGTTCACCT AGAGCTTGG CAGACTCTG CCCAGGCGAG GACTGAGGCA 660
AGCTCTAAG CACTTCTAGG AGTGTCTCT TCTCAACAG ATGAACTGAC TGTTTTCTTG 720
CGACTACTG CTTTTCTCT GTGCCACCA CTTTGGGAG CATTAGAAA AGGTGCTCT 780
TGTGGGAT TCTAGACCA CAGGCCAGA CTTAGATATC CTGGCTCTC TGGCCGCGG 840
GGAGGAGAG CTGCTCTGCA CCGAGAGAA GCGACTGCT ACTGCCAGG TGAACCTCG 900
GGGAGCTCG CTGTCTCCG CCGCCGAGAG CTCCGGAGC CCGCAGAGC CGGCGCTCT 960
GGCCCCCAC AGCCGCGAGA TCCCGGACC CAGAGGCGCG GTCTGTCTC AGCCGAGAA 1020
GGACTGCGC AACTACACT GGAATCTCTT CGGCTGCGC TTGCGCAGC GCGAGCGCG 1080
ACAGGGAAC CAGCGCGAA GCGTGGCGG GCGCTGAGG CCGAGGTGCG GCGCAATGAA 1140
CTTCAGACC CAAGGCTC AGAGCATCG GCGCGCGCG GCGCGCGCG GACTGAGCG 1200
TAAAGGAGG GCGCTGAG CTTCACCC GAGCGAATTA AAGAAATTT GGTAACTCA 1260
AAAAAAAA AAAAAAAAAA TCGCGCGCG 1290

(2) INFORMATION FOR SEQ ID NO: 82:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TTTATTTGAT TCTTTACTA TGAATCTCT AATTAAATCT TTTTGGACT TCTTAAGTTG 60
TCTTTAAGG TTTTAAGTC CATCTGAG TTTCAATAT TCACTATCC CTTTAACAT 120
GAGTTTATT ATGACTCTT ATGATTCAA AATCTACAT CTTTTGTAG TCTCTTTAT 180
TTTTCACCTG TTCTCTTGA TCTTACTCA TGTATTTTA ATCTCTGTT TTTTTFCTC 240
TTTAAAGA CATCTTTTA AATTAATTT GAGGAATAT TTGATCTTA TGAACAGCC 300
ATTACTCAC AGAGAGATT TTTTGTGTT ACAGTCCC TARGAATCT ACAATCTCG 360

420 GAGCAGTAG AACACAGGT GATGAGACA TCTTGSGAT CTTGTTTAC TTTCGSCAT
480 CTTTCTCC AACCTCTGG GAATATCAT YCATATCTTA RTGAGGCT ABAAGTGCT
540 TTATCAGAGC CCACTTCCA GGGCTCTGG CTTTACTAC TGTACCCCA TCATACTGA
600 GCTCAGGA TTGATCTCT TTTTACTTT CAGATTTCT TTTAAMATC TTTTATTTT
660 TTTTCTTCC GAAGATTC CCAACATTA CCATTCCCA CTTTCGTTG AATTTTCTG
684 GCTCTCATTT TGATTTTTC AAGA

5
10
15
20
25
30
35
40
45
50
55
60

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2024 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

60 CTCCAGCAAT TGGCAGAG TGGCTGGAG GCTTCACTTT TCCGCGCT GCGTCCGCT
120 TCTTGGAAT GGAATCTGA GATTCTCTG TTGTTCTGTC GCGGGTTTC GCGCTTCTC
180 GCGCTCTGG GCTGCGAGC TGGGAGGG GTTGGAGGG GCTGTGATC GCGCGTTTA
240 AGTTGCGTC GGGGGGCGA TGTGGCGCG GAGAGTCGAG GCGCTAGTGT GCGAGCTGAG
300 CCGCGGAGC GAGGGGATG AGGAGGAGA GTGGCTCAT TGGGATGAA ATGAGTTGA
360 AAGGCGAAA GAGAAATG CAGATCTTA TCTTCATCT GGAATGAG ATGAAACTGC
420 TGAAATGCT GTACAAAAC CGAAATGAC TGAGCTGGA GATGATGAG ATAGTGACAG
480 CGATGATAT GAGATGATG TTCAATGAC PATAGAGAG ATTAAMAGG GAGCAGACA
540 GTATGCGAT TATGCTACAG CACTGTAAA TCTTACATC AAGCAGGGG GAGAGTTTA
600 TGGAACTACA GGGCAAMAG TCAAGAGT AGACTTCAT GCACTCGAA GATTAATGG
660 AGTTCCACT TTAGAGTAG ATTTGGATC TTTTGAAGT AAACCTGCG GTAACTGCG
720 TCTGATCTT TGTATTAAT TTAAATGAG GTTTATGAA GATTACTGA AGCTTACTG
780 TGAAAMAAA AAGAGTATC GAATGGACT TGAAATTA TA CAGTAAGT CTACTACAA
840 TAAATTTAG GTACAGAGG GAGAGCTGG AAATCAGAG AAGAAACTG CCGTTCTGC
900 TACAAAGCT GAGTTTACT CTTCTCTTC TTGTTCTAG ACTGGGCTC CACCGAGAG
960 GAGTTTACT GGGGAAATG ATGTTATGG TGAGACTATA ACTATCAGCC GATTAAGAG
1020 CAGGCGAG GCAATGAGA ACGAGCAT AGAGTCTCT TGTGAAGAT CTGCTACTGA
1080 AATGAGAAC AATTTAGCA AACACTTC GTTTTCTCT CAGAGCTC CTCGACTGA

1140 CTTTCAGCT CTTCAATTC TTCACTGCG TGGAGTGTG AGCACTGCT CACTCTCAT
1200 TCCACAGG GATTTCCTC CTCACAGG GCTTCACT CCACTCTTTA TACAMCAT
1260 AAGAGTGA CATCTCTCT GTTATGATG TGTCTCTGA CTTCAATTC CATATGCGA
1320 TTTTCTCTT CCACTCTTC CTGCTCTGC TCTTGTGCG CTTAGTCTTG TGGAGAGAG
1380 CAGCAGTGG GACTATATG CAGAGAGCA GAAAGAGCA GATAGAGAG GAGCAGAGA
1440 CAGAGAGCA GACCTGATC GGGAGAGCA AGAGAGAGC AGCAGAGAG GAGAGAGGA
1500 GCTGATGAC AGTCTTACAC CAGTGTCTT CAGAGAGAT GAGAGAGAT AGAGTACAG
1560 GGAATATGA GAAAGATTT ATGAGCTCA CAGAGAGAT CAGAGAGAT AGAGTACAG
1620 TACAGAGAG CAGCAGAGG AGAAAGAGCA AACAGAGAT AGTCTCTCT GAGTATATG
1680 TAGAGTGGC CATGAAATG AGAGAGAGC TACTCAGAG AGAGAGAGC AAAAAATTC
1740 TAAAGAGC AAGAGAGAA AAGAGAGAG CAGTACAGCT CCGCTGAGC AGAGAGAGC
1800 CGAGTACCA CTTGAGAGT AGGAGAGAT TGGGCTTTT GTTATATTA GTACAGAG
1860 TACATATAT AATCTCTTT ATTCTCTCG ATATCTTTA AAAAAATTC CTTAATCTT
1920 GTCTCTTTG TTAGTATGAA AGTTTACTT TTTTTCAAA ATTAAGAGT GATTTTCTA
1980 TGTATATTA AAAATCTTG TCTTACTTA TTTCAAAAT AAAGAGAG CAGTACTTT
2024 ATATCAAAA AAAAAAANA AAAAAAGC GCGC

5
10
15
20
25
30
35
40
45
50
55
60

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 931 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

60 CCGCGATA CCGGAGCGG GATCTAGCT GCGAGATGA ATGTGGGCT GCGACAGAG
120 GAGTAAAGC CAGCAGCGG AGTATGAT ACCGAGGCA TCTGGCTGC CTACATATC
180 TTGATAGAT TCGTCAATAT GGTCTACTC AGCATCTCT TCTTAGCAT TCTGTGTTT
240 TGGAGCTGA CCAAGTCAAT CATACTGCT GCTATGATG TCTTCTTTA TACGATGAA
300 GCGAGAGCT TTGAGAGCTG TGAGAGGCA AAGCTGGGC TACTGAGACA CTGGAGGCA
360 ATGAGTATG GCGTCAATTT TACTCTTCC GCGAGTTCC TGAGATCTC TCTATTTTG
420 CTTATCTCC TGGCAGCTT CTATACAGG TATGATGCT GCACTTCTT CATCAGACA

5 GGCCTCATTC TAAGTGTACT GCTCCGGAAG TTCCCCCACT TCCATGGGGT TCGTCTCTTT 480
GGCCTCAACA AATACTGAGG GATGGTGTIT GGGACAGCTC CHTGGGCATG GGGAGGCAAC 540
TGAACAGAG GACTATATAA CATCTCTCTC TTATTTCTCA TACTGTCTTC TACACTTTTA 600
AAGCTCGAG ACTATACAC CTTTCCGAGA CTTCCGAGAA GAGAGAGAT TTCCAAATAG 660
GGCTCTGGG CCCAGTCTTG CTAGTGGCAA GTTCTTTTGA ATCAGGAAGG CAGTGAAGT 720
AAGGCCGAAA TCACTCTCTT CATTACGAGG AAGCCATTTG GGCAGCTCTT TTGGTGATTA 780
CATCTTTCCA TACTCTTTAC ACTTACACCC TTCCAGCTCT GTTTTGGCTGT GTATTTTCT 840
TCAATTAAT TTTTTCAGCT ATAGCTCCAG TTATATCAGG ATGGGTATAG AGCTGTCTTC 900
ATAAGGCTGG GGTGGGAAG ATGGATTAAT G 931

20

(2) INFORMATION FOR SEQ ID NO: 85:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 825 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CCGGGCCCGC GGGGTCTTCA GGTACCGGG CTGGTTACAG CAGCTCTACC CCTCAGAGG 60
CAACATGCG ACCCGAAG GACGACGAGA AGATGCGGA GGGCGAGGG CTGAGCGGCA 120
CGACCTCTCT GCGGAGCTG ATTCCCTCCG GTCCAGGCCG GAGTGGCTG GAGCGCGGCC 180
GCGCGACCAT CCGGCCCTGG AGCACTTCCG TGGACGACGA GCGCTTCTCA CCGGCCCGCA 240
ACCTGGGAGA GCTGTGGCAG GCGCTGTATC GCAAGGTGGA GTACTACAG AGCACTATG 300
TGTTCGTGTT CTTGGGCTC ATCTGTACT GTTGGTGAC GTCCCTATG TTGCTGGTGG 360
CTCTGGCTGT CTTTGTGGC GCTGTATTCA TTCTATATCT GCGACCTTG GAGTCCAGC 420
TTGTGCTCTT TGGCGGAGAG GTGAGCCCGG CGCATCAGTA TGTCTTGGCT GAGGCGATCT 480
CCTTCCCTTT CTTCTGGCTG GCTGTGGCGG GCTCGGCCCT CTTCTGGTG CTGGAGGCA 540
CCCTGTGGCT CATGGGTC CAGCTGGCT TCCACAGAT TGAAGCTGTG GAGCGGAGG 600
AGCTCGAGAT GGAACCTG TGAGGTGTCT TCTGGAGCT GCGGCCCTCC CCGGCCAGCT 660
GCCCAACCCC TGGCCATGCC TGTCTGTGAC GCTCTGTGCT GTGGGCCCA CAGCGCGGTC 720
CATTCAGAG CCGCGGAGG GATCCCGGCT TTGAATATA AGCTTTATG GTTGTATTC 780
AGGAAAAA AAAAAAAG GGGGCCCTC TAGGGTCAT AGTTA 825

60

(1) SEQUENCE CHARACTERISTICS:

(2) INFORMATION FOR SEQ ID NO: 87:

(2) INFORMATION FOR SEQ ID NO: 86:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

10

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

CATGTAAAG GATGAATAT GACTCTGCT GTTTTTTTAT TTCTATGAG GCACTTCTG 60
GGAGCGTTT CTGGCTCTCA GGTCTGTAGA AGCTCAGTT TATGATGTC TCTGTGTGT 120
CTGCCACTA CTGGAGAGC CATAGCTCC AGCTTTAGA AAGGGAJCC CCGGGCAGAG 180
TGTGGGAGG TGGATGCA GATGGGAGG GCTTTGGAA ATGAGAGGTG AGATATATC 240
CAGAGAGGT GTAGAGAGG GATGGAATCT GATACATGA TTAGGGTCT 300
GTCTGGACA CTGGCTTCC TGGTTACTG CTCTTCTCT CTCTCTTGT CCGAGCAGG 360
GCTGGCTAC TGTCTGTGCT TCAATTTCCA GAGCTGGCTG CTGCAATCAG ACTTATGTC 420
TCTTCTCTCA CTTTCTCTCT TTTCGGGAT AGTGGAGGTG ACGAGATGT GATTTGGGCA 480
GGATGTCT TTGATGCAAT CAGACTTTA GCTTCTGTT GCTGTGTGTC CAGCTCTGAT 540
TTGATTTCA GCGGTGATG AATGTTGCA TGGAGCTGA GACTCTACT GAGCTGAAA 600
CCTCAAAATG AACATATCC CTGCTTCTCT GCGAAGGATC CTTGTAGGCT ATCCCGAGCT 660
TCCCACTCT TTTTCTGTGT CTGTGACAG AAGACAGAG TACTTGTAT GCGCTGTGAC 720
CTGGCCAGTT GATTTTCCC TCGAGGCTTG AGCCAGGCC AGAGCTTTGA AAGGTATTC 780
AGGTTTTCG CCAAAACAT GAAAAAAT GCGCTGGGCC TGAACCAAT AGCTTGACC 840
CTGTAAACT CATTAGCTG ACCCTCTCT TTGTGATATA CCGAGGTAGA ACAACTCTCT 900
CTCACTGTCT GTTGTGAGA TACGCTGTAG CCGACTCAT TATGATATTC TCTTATATA 960
TGTTTTGGAC TGAATACCT GCGATCTTT TTCTTGGGC AATCTATAT TTTTCTAGA 1020
GTTTCCAG GCTTACTGA GGAATTAAC ATACTCTTAA TGGCTTTCT CTCTCTTGT 1080
TTACTTATG CCGTCACTTC CTGATTTTAC CTTCCAAATA CAGGATTCAC CTGTACCCA 1140
GCGCTTACT TCAAGTATC AGATCACTCT GTACCGAGC CTTTACTCA AGCTCTGCTT 1200
TGGAGAAC CAATATGA CAGTCTCTCT GTTGGCTT 1238

55

60

- (A) LENGTH: 1460 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

```
5  ATTGCTTTCT GATCTCTGCT GACACTGGGG TCATCTTTCA TCCCTGGAGA GCATTTCTGG 60
10  CTGCTCTCTC TCACCTGGGG CTTGCTGGGG GTGCGGGAGG CCACTTTATC CACCATCGGG 120
    CCACTCTCTA TTGCGACT CTTTTGTGGC GACACGGCG ACCGATGCT CAGCATCTTC 180
15  TACTTTGGCA TTCTGTTGGG CAGTGTCTGG GCTTACATTTG CAGGCTGCCA AGTGAAAGAT 240
    ATGGCTGGAG ACTGCTACTG GCTCTCTGAG GTGACACGG GTCTAGAGAT GTGCGCTTT 300
    CTGCTGCTGT TCTCTGTATG GCGGACGGG CCAAGGGGAG CCGTGGAGCG CCACTGAGAT 360
20  TTGCTACCCC TGAACCCGAC CTCTGTTGGG GCATATCTGA GGGCTCTGGC AAGAAATCTT 420
    AGTTTGTGTC TGTCTTCTCT GGGCTTCTCT GCTGTGCTCT TTGTGACGGG CTGCTGTGCT 480
25  CTGTGAGCTC GCGCATTTCT GCTGCTTTCC GCGTGTGTTG TTGGGGAGAC CCAACCTCTC 540
    CTGTCGGGAG ACTGCTGCTC TTCTCTCTGAC AGTCTGATCT TTGACTGCTAT CACTGTGCTG 600
    ACGGAGATTC TGCGTGTGGG CTTGCTGTGG GAGATGACCC GCGCGCTTGG CCACTGACAC 660
30  CCGCGGGGCT ATCCCTGTGT CTGTGCTGCT GCGTCTGCTG GCTCTGCTAC CTTCCTCTTC 720
    CTGTCTCTTG CTTGCGGGGG TGCTAGGATC GTGGGCACTT ATATTTTCTA CTTGATTTGA 780
35  GAGAACCTTC TGTGCAATGA CTGGGCAATC GTGGCGGACA TTCTGCTCTA CCGTGTGATC 840
    CCTAACCGAC GCTGACCGCC CAGAGGCTTC CAGATCTGTC TGTGCTGACT GCTGCGTGTAT 900
    GCTGCGAGCC CCACTGCTAT TGCGCTGATC TCTGACCGCC TGCGCGGGA CTGCGCGCCC 960
40  TCTCTCTGTT CCGATTTGCG GGTCTCTGAG TTCTGCTCTA TGCTCTGCCC GTTTGTTGGG 1020
    GCATGCGGCG GCGCACTTCC TGCGACCGCC CATCTCTATT GAGCGCGACC GCGCGCGGCC 1080
45  ACGCTGTGAC GTGAGGGGCC TGCTGCACTA AGCAGGTTCC ACAGACGACC GGAATTTGAT 1140
    GCGCCAGCGG GCGCGCTCCA CCGCGTGGCC CCGTGGCAAT GTGCTGCTCT GAGAGGCTG 1200
    CCGCTGACTT ACTGTGCAAT CTGCACTGAC TGCGCTTGGG CCAACCTGAC GAGAGGCTG 1260
50  GCGCTTAAGC CTTTGGGCTG GCGCACTTC CAGAGGACC CTGGCGCTTG TGCGAGCTCC 1320
    CAGACATCAC ATGCTGATGT CAGGCGAGGA GCTGGGGTTC CAGGAGGGGG ATCCCTCTCC 1380
55  AACAGGGCCA GCGCCAGGCG CTGCTGCTTA TTTGTGACGG GATTGAAAT TTGAGCGAGA 1440
    AAAAAAAAAA AAAAAAAAAA
```

(2) INFORMATION FOR SEQ ID NO: 88:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1395 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

```
5  CAGGTGCAAA GTGGAAATGG TCAGTCTCTA GTCTGGGCT ATTCGGCCAC GTGCTTGGG 60
10  GACATGGGAC GCTGGAGGCT CAGCAGCTGG GATGCTCTGG CTTTGTGCTC CAGCGGTGGG 120
15  AAATTTGGCA TTGCAAGGCG GGAAGCTGGG ACTGAGCTGG CCGCCCGGAC GTTTCCTATC 180
    CTTCCACCGG ATTCGTGGGC GCTGCGACTG GCGTGTATGT AGTTTCTCTA CTTCTGAGCC 240
    GTATTTCTTC CAGATTAAAG GTACGACATT TGAGAGGCCC AGCGAGAAC GTCAACCGCA 300
    GAAGCTTCAC CCGCGGAGAG CGACCCGCTT GTGTGCTGCC CCGAGAGGAC AGCGAGTTTG 360
    TAGGGGGGAG TCCCACTCTA AAAAAAAAAA TCCAGTCTCC CAAGGGTGA CCGTCTTTCG 420
25  GAGACGCGG ATCCGACTAC ATGTGCTGTC CCGAAGAAAT TTGACTTTTG ACTGCTGAC 480
    TGCTGACCCC GGGTGTATTT CAGAGAGAA GCACTCTCTC CTGCTTGGAA GAGACTTCAC 540
    ACGCTATCA CAGTGGCAC GGTCTGTGAG GTGATGCGA TTCTGCTGTT GATTCATCAC 500
    TCCGCAATCA AAGAGCCAA CAGAGCCCAA CTAGAAATAT GGTTCGCCAG GGTCTGGTCA 660
    GCGCCCTTAA AACTGCACTT AGTTGCGGTG AAGCCTTAG ATTAATCTTT TTTCTTAATT 720
35  TTTTAAACCA ATGCTATGCT TCTGTCACTT TATGTATCTT AAGACTCAAT ATAAACCCCT 780
    TTTTAACT GAGGAAATCA ATGATTTGAT TCCCAAAAAA CAGAGTGGG GATGTATATG 840
    TCGAACCTGG TTTTACTTAA CCGTGTTTT AGACTTTTTC TTTCCTTTTAA TCACTCAGCC 900
40  TTGTTTCAC CTGATTTGAC TGTGCTTAG CTAGAGGCG CAGATGCACT CCACTTGGC 960
    TCTTTCACT GCGAGCGGCT TCTTCAAGG ACTTAACTTG TGCAAGCTCA CTCCGAGCAC 1020
45  ATCCAGATAT GCATTTACT GATTAAGTAC TGTGCGAGCC GCGCTCTATC ACCTTTATAT 1080
    ATTCGTCCA TTGATTTAC CCAAGAGCCC GCGCTCTATC ACCTTTATAT 1140
50  CCGCTGACC TGGAATATTT AAGCTTCTTG TAACTTTTAA TCTTTTAAAC TTTTGTGCT 1200
    ACTTTATTTT TGTAAATATG TTTTAACTAG ACCCGCCCTC TCGTTTCTTAA ACCAAATAT 1260
    AAAGCAAT CTAGCCCTTT CTTCAGGCGG AAGAGATTTT GAGCGTTAGC CTTCTCTGG 1320
55  CACAGGCTA ATTAAGCGA TTCTTCTATG GTAAAAAA AAAAAAAAAA CTCCGAGGG 1380
    GCGCCGATTA CCGA
```

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1186 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

5 GGCAGAGCC GGCAGAGCCA GCTAGAGTCA AACTGCGGG GGCAGAGCA TGTNACAG 60
AAMACAGAA GATGAGACTC TTTTCAATCA GTTTTCCTAG GGCATCTCG TGTTCATCTT 120
TCCCTCTCC ATCATACCTC CTCCTTCTCG GAGCTCTCG CGGCTTGCT GTANTGTGG 180
CACTTACCTG GATTTTTCAG TGGAGGAGTG AAGGCGAGA CTCACCTTAC GCGGTGGAC 240
AGATGGGAG AGGAAAGAG GAGAGATGCG CAGGAGGCG GTGAGAGCA AACGAGAG 300
GTTGGGTGAG GCGAAAGGCG TGGGAGAGAA GAGGGGTGCA GGGCTCTGAG GCGGTTTAC 360
CAGCAGCTGC GGGCTCTGCG GGGCTTGGG ATCGAATCTC GAGAGAGCG TACGAGCTC 420
CTGGTGTGTA TCAAGAGTT TTTTCAATCA GTTTTATGCA TGAATCTCT AGGTTTACA 480
AGCGCTGCG GTGAGCATG ACATCCAGC GATATATTA AGATATATA TATATATTA 540
TGTATATATA TTTTGGGCG AGAATATTA GCACTGAGCG CCGTCTGCG CTCGTGGCC 600
AAGCAAACT AAGCTCTTTC GTTTGGGAT TATGTTTCT TTTGTTATT GTTTGTTT 660
GTGGCTGCG TTTTGTGCG ATACGACAG TCGAGTGGG ATTGCTCTGT ATTACAGAT 720
AGTGTGTTTA ATTCAATAT GTTTAGTGA ATGTCTACT CAGCAGCTCC TCTTAGCTTA 780
ATTTTAGAG GTTGGCCAT TTTGTTTCT GATTTTACT GGTACTTCT TTTTACAAAT 840
CAATCTCTT CTCCTCTCT CTCCTGCGA CTCCTGACC CTCCTGCTC CATCTGCTC 900
TCCGCGCTC CCGTCTGCG TCTGCTGCC GGTCTCATTT CTCCTGACT CATCTCTCT 960
CGCTCTGCG TCGCTCTGCG TCGCGCTGCG CAGGCGACT TCGCGGAGT GTGCTTGGG 1020
CTGCTTACT GTTCTAGTTC CGAAGCAATT TCACTGAGAG TTTGCTGAG CTGCTGCG 1080
CTTTCTGCA CTCGCTCTCT TTTCTGTAGA TTGACGGTT TCTTTGTAT TTAGCTGTT 1140
CTGACAGAT TTAAGAGAA AAGAGGAAA AAAAAAAAA AAAAA 1186

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1821 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

5 AAMACATGCT TTCAAGGCGT CCCATATGTA TTGCGGGGCG CAGAGACAC TCAGGCTGCA 60
KATCTGCTCT CAGTGGCTTC AAGATGCTCT CAGGAGACAC CAATTAACCA GCTGAAATTC 120
ACATATGCTC CCAGAGTCTT CCCAAGCTTA GTCTGACCA GAGGTGGGCG TCTCAGGCTA 180
GAGGTTTTCAC AACATATGAC AGGCTGCTCG GGGACATTCG AAGAGCCCTT TTCTTTGCT 240
CTCATGCTTA GAAGCGAGCC CTAGAGACT CAGTTACTC CTTGTACTC AGCAGAGAGC 300
TGATTTGACA CAGCTGGCCA CAGAAAGCCA GTGGTAATAC ATCTGTTTAC GTTTCCCTAT 360
CAGCCAGACA CAGGCCCTTT TCCAGGTGCA AACGAGAGC GATTCATCT CAGTTTTCAC 420
AATCAAAATCA CTACTTCCAT TCTTACTTTA GATCAGCCAA AGTGTGACT GCTGCACTGT 480
GTGCTATGTC CTACAGGCC CAGCCAGCG ATGCCAAG AGCCAGCTTC TCCAGGCGCTG 540
CAGCAGAGCG AACCCAGCA GCTTAATGTC AGCAGAGAC CTCGCCACCA ATGTCTTGT 600
CTAATTAGAA GGGAGATTG GGCACAGAA ATCACTTAT CTATATATAC AATATCTCT 660
TGATCACTT TAAAGTCTCT ATTGATCTCT ACTGCTTTTA AAGCTATTTG AAAAACTCA 720
TACTAATACA ATGACACTC TAGAAGATT TGGAGGCGC ATGCTGAGAA CAGTTCTCT 780
GAGGTGAGAA TTTTTCGAG AGCTACTTAC CTACATGTA ATGTGCAAT TTCTTTTCTT 840
TTGTAGAGAA AATCGCTTT ACTTTTGGC AGATTAATGG CAGCTTCTAG TACAGCCATT 900
ACAGTTTCT ATGAGAGAA TTAAGATTA CTATTAATTT GTTAATATAT CAAATATGG 960
ATAATGATGG CCAGAGATTT TACATACAA AGTAATCTC AATGTAAGC TATTCAGCTC 1020
TTCAAGTTTG AATCGCTCT AAGCCAGCTT GAGCTTCCAC ATCATCTTCA AAAAAAGTT 1080
TCTCTGTGTC CCATCATCT CTATAGGTT AACTGTTTAT TCGTCCATTT AGCAGCTTTT 1140
AATCTGCA AAGATATAT ATCATGTAAT TATTTTACTT TTACAAAAA AAGTTGTAAT 1200
GCGGTTTAT TTTCATGCGC TATAGCAGG TACCTTAGTA GGGCAGATAT AGGAAAGCA 1260
AATGAGACA AAGCAATCC TCTAAGATC CAGGCGAGA AAGTGTGTGG CAGATGACT 1320
CATCTCTGTC TCGCTCCCAT CAGTCCAAAT CAGAGGCGTG CAGTGAATGC CTGTTCTTTG 1380
AATGTGAGC AGTGTGCTCT GTACTCTTT AAGACTTGGC TATAGGCTGT TTAGCAAGT 1440
ACAGATTAAA GATACAGTTA CTTAAACAGC AAGTAATTT TATATGCTCT CATCCATTTA 1500
TCACTCTTTG GTTTGCTAAT TTTTTCATCT ACTTTTCTCT ATCAGATCT GTTCTTTTGG 1560
TACATTTTC TCATATTTGG GTTGCACAG TAAACACAAA CTGCTATTTTC AATGAGAAAA 1620
GTTATCTGTA TGGATATTTA AACCAATTA ATGTTATAAA GGGTAAAAA AAAAAAAA 1680

AAAAAAAAAAAAAAAAAAAAAATTC CTCGCGCGCG CAGCTTTT CCGTTTGGGT 1740
GAGGGTTTAT TTTGGCTTGG GCGACTGGGC CTTTCTGTTT TACAGCTTG TGAGGGGGGG 1800
AAGCCGGGGG GCGTTTCCCC C 1821

5
10
15
20
25
30
35
40
45
50

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TCGCCCCCTTTT CCGACCGATT CCGGGGCTGG TCAAGGTGGG AGATGTGAGC TCCAAATTAG 60
GAGCTGGAGA GAGGTGAGA ATTGTGAGG TGGGAGATT GATTGTGAT CTGACTTGT 120
AATGAGTGT ACCTTGGCAT CTGTGTTCAA GTTCACGGTT TGTGTGTGGG TTCTGTGGAG 180
AGCTTACTCA CCGCGAGTC TTTTCTTCT CTCTCTCAA GAGAGCCCT GTTGTGCTTT 240
TACAGCCGT TGAATCTCT CAGAGACAA AAGAGCAGA CAGGGAGCTG TAGGGAGAT 300
TCTTTCTGT TTTCTGTGT TGTCTTTTA CAGGATGCC GAGAGCCAC TCATGGCAT 360
GCGAGAGCT TTCTCAGAA CAGTCAATAA CAGATCTCTG AGTCTCTTTC TTGTCTCCC 420
AGCTGAGCTT TCTTATTTCA CCGTTCTGG TGTCTATAGG AATGCAATAG AAGACCTTGG 480
GAGCTTTTTC TGTCTCTTTC TGGCCCTTCA TGGAGCCATG GCGCTGGGCC TCGGGGAGTC 540
CTCAACCTCA CATTATTTT CTTCTCTGG TCCAGCCCT TTTTGTGTGT CTGAAACGGG 600
TTTAAATATG TGAATCTCCC AAGAGAGAG CCGCTGGCTG TATGAAACTT GAGCGGCTT 660
TTGTAAAGTG CCACTCCCAA ACTTAAAGT AGCTTAAACA ATTTTTAAAT GATTCAATGG 720
CTTGTTCATC CTCAGATGT AGCTATTTGAT GTTACTTTGG CAGCGGAGTG TCTGAATTTG 780
TGTGTGCTCT GATTATTAGG ATTCAATAT TAAATGTCT GCTGAATATA AAAAAAAAAA 840
AAMACTCA GGGGGGGGGG GT 862

(2) INFORMATION FOR SEQ ID NO: 92:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 696 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

CTGAGCGAG TGAATGGAC TGTAGGGCT ACCGCTAGCG CCACTGCTGC GCGAGGGGG 60
TCGAGGGGAG AGGGGGGGGG AGGGGGGGAT TGCMAAGATG GCTCAGAGCA GAGAGGGGG 120
AAGCCCTTTC GCGAGGCCA GCGAGCTTGA CAACCCCTTT CAGGAGCCAG CTGTGATCCA 180
GCGAGGAGCC AGCGGGGAT ATGCGAGCT TGAATCTTAC AAGCTTTTTC AGAGCCGGGA 240
GCGAGGAGCA GCGTATGAGC CTGAGAGCC TGGCCCATTT CCTCAGAGCT CAGCTTCTTC 300
CTTTCAGCCC TCGAGAAAGC TGAAGGCCAC AAGACCTTAG AACTATGCTT CATACAGAC 360
TCAGGGCTCA GCTCAGAGAG CCGAGCTTGA GCTGCTGAG AAGCAGAGG AGCTCAACCG 420
GAGGGCAGAG GAGTTGAGCC GAGGAGGCA GAGCTGAGC ATGCTGCCCT GGGGGGACAA 480
GCTACTGAGC AGAAGATTTG GCGCCCTCTA CTTCTTTTTC GTTCAGTTCA GCGCTGCTTT 540
TTTCAGAGCA TCTCATGGA GATGCCCAA GAATTTGAGA AGACTGTATC CACCATGTAC 600
TACCTCTGGA TGTGAGGAC GGTGATCTT CTCTGATTT TGTGGSCTG CCTGCGCAGT 660
TCTGTGTGGA AACCAATAT GCGAGGCTT TGGTT 696

(2) INFORMATION FOR SEQ ID NO: 93:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1886 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

CAGGGCACTG AGCTTCTTT CCGAGGGATG CAGGAGTCC TACAGAGAAA GCGCTTCTT 60
GCAATTTAGA GGGGGGAGAG CTTGTGAGC AGAATGACC AAGCAGCTTT CTACTGGCT 120
CTGAGAGCTG CCACTGCGAG AGAGATCCA GAGGCTGTGG GTTATGTGG CCAAGCTCTT 180
CAGCTTCAG GTAGCAATGC CAATCTCCCT CACTCTCTTT CCGCTGCTGT GTGAGCAG 240
AAGCTTACC ATGAGGCTCT GACATCTTC GACATGGCCC TGAATGATA CCGAGAAAT 300
TTCTACTAC TTTTTCGAA AGTGAAGTTG CAGTCACTCT GCGAGGCCC GAGAGAGCA 360
CTGCTGACTT GTAGGAGAT CTTGAGGATA TGGAAATCTT GCTACAACTT CACCAAGCCC 420
AGTGAATCTG GAGTTGGGAG CAGCTCTTTA GATAGAGCA TTGCTGAGAG AGCAGAGCTT 480
AATGAGATTA CTTTCCGAGA CTTGAGGAT CCGAGAGAG GCTTCCCTCA TCCCAATGG 540
GTAGAGAGCT CAGAGATGGA GAGGAGCTG TCGAGAGTGG CTTGCTCTCT GCGAGAGATG 600
CGCTTAAACA GGGGGGGGCT CAGCCCTTGA TGAAGCTGGC ACAGATCTGG CTGATGAGC 660

CTGAGTCTTA TTGCGGATC GGGAGCGCTG CAGAGGCCAC AGCCTGTACC CAGAGAGCTG 720
CGAACCTCTT CCAGATGTC CAGATGTGTC TGTACATGCG CGGCCAGATT GCTGAGCTGC 780
GGGAGAGCAT GAGAGAGGCG CGGCGGTGGT ATGAAAGGCG CTTHAGGCAT CAGGCCGACC 840
CAGCTGAGAA GCGATGCGCG ACTTGCGCTT GATCTGTAC CAGTATGCGC GTTACAGATT 900
GGCGGAGAG AGCTCTCGGG AGCGGTGCGA GGTGAGCTCG AGAGGCCGAG AGGTCTGAAA 960
CGGCTGCGGC GAGGTCTGTC AGCTCAGCG CAGAGATGCG CGGCTACGCG AGTGTCTTCT 1020
GAGAGCGCTG GAGGTGAGCG CAGAGAGGCC CGCGGTGCGC TTACAGATCA TCCCGCGGCT 1080
GCTCTGAGCA GCGCGCTGCG AGCCTCACTT GCGCTCAGC CTGCGAGGCG CTTGCGCGGC 1140
ACGAGGCTTT GTGCGATGCG CCGAGAGGCC TGAATCTGCG GAGCTGAGGC CAGGGAGCAG 1200
TGTTCAGTGG GCGAGATGCA ACCAGCGAA CCGAGCGCGA ATCACTGCTC TCGCGATGCG 1260
GTTTCTCTTT GTTTTCTTTC CCGAGCGCAT GGTATGTTCT GAGCTATGTC ACGTTGTCTA 1320
AATGCGATCA TGTGCGATAT TTTGTTAGTT GAGATCTGAG TTTTCAGTAA AATGATATG 1380
GATTAATCA GCAATGTAG AGAATATAT TGAAGTTAA AATTCAGTGG CAGCAGAGAT 1440
TATTTTATTC AGAGCTGTA AGAATACAC TGTCTTTTC TCGCGAGCAC CGTCTGCGC 1500
CGACTTTGCG CCGAGAGCA AATGTGACT TCGTGTCTCC CAGCTCAGCA CTATGCGATG 1560
CGAGGAGCAC AGCTGAGCAT TTTCTGTTT TACTGTGCAT AATGTGACCA TGTGATCAT 1620
TACTGTCTTC ACTTACAGCA AGCGTGAT CCGAGATAT TTAATTTTA CCAATATATG 1680
CGCTGTACAA GAGAGGAAA TATGAGTAT TTAGTTTAA CTTTATATG TGAATTCAGA 1740
GTTTATTTAT CAGGCGAAT ATGTACAAAG AGCTTCAAA TGGATATATT ACGGACATTC 1800
CTTATACATG ACAGACACTT GCTACATGCG GAGATGATG TTAATATTA AATGATTTT 1860
AATGCGAAA AAAAAAAAAA AAAAA 1886

(2) INFORMATION FOR SEQ ID NO: 94:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1774 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

CTGAGTCTTA GATACAGTA GAGATACACC CAGTTTCA TGCAGTACAG TGGAGCTTGC 60
CTGCTCTGCC CCGACATGCA AGATGTTCTT TTTTCAATAC TATATACTAT TGCATTCAT 120

GATATCTGCG TAAATTAATT AATCCCTCAT GGAATGATA TTAGCTTCAT TATAGATAGA 180
AGCTATATTA ACATTCCTGT ACATGTATTT TGTACTGTGT GTGGTAATTT CTGTAGATG 240
AATAGTACA AATTAATGCG ATCAGGTTTC ACATTTGAG TTTTGAANAC TACTAGCGAA 300
AAGATTTCAC CAAATTACAA CTCATCAT ATGAAAGATG AGTGTGCGC CTGTGTTGCG TATAGTCTGC 360
CAAGCTCGAA TCGTTAANAA TTTTTCGCGA TCTGTGAGGC AAGATTCTCT TCTTTTCTTT 420
GAATATTAAT GAGGAGGAGC ATCTTTTCAT GTTTCCTGCG CATTGCGATT TCGTATATAG 480
AATGCTTTT GCGATTTTC CTTTFTTAA TTATGAACT CTATGAGTA CTTTCTCAT 540
GTATTAANAA CAGATCTCT TGAATAGAG GAGCCTTTTC TCGATGCTA CCAATCATAT 600
TCCACTTACC ACAGTTTAC ATACATCTCT TACTGACTTT TCGTACGAA TATACATACA 660
CATTAANACA CTTTATACAT AATAGGATC TCAATCTG TAGCTTTTAA AATTTTGTGT 720
CTCAANAAA CATTACAGCT CTTTAATTT CTTTAATGCT TGAATATAT TAAATATCAT 780
GAATATGCG TTAATATTC CTTAATTTT TTTCTCTCG CTATACAT TCCANAGTAA 840
ACATCTCAT CAGATCTCT TGTGATGCG TGTGAATTT TCTTTAGCT GAGTGTGAT 900
AAGGTGATTT TTTGATCAA AGGTTTGT TCTGTGCGC CTTCACTCT CCGAAGGCGC 960
TTCAATAGTG TATTTTCAGC AGTGTAGCG AGATGTGCA TTTCCGATA TACGATACCC 1020
TACATCTGAT AGTTTATTC TGTTCGCGCA AAGAAAGCT TTTCTTATTT TCGATTTGCC 1080
TGAATTAATA AAAAAATGCT GAGATGCGG TTAATTTTCT GTTATTTGCG CATTATATAT 1140
TTACTGTGGA TGTTTGTAT CCGTTACCTG CTTTCTATG GATTATATGT GATATATATG 1200
TTTATATGTT TTCAGATCT CTTTCCCAT CTTCTGCTAA CAGAACTTTT ATTATTTTGT 1260
GGGAACTTA TTTCTGTGCG CTTAGGTGAG CATGTGAGCA GCGCTGCGCT CCGAGATGCC 1320
ACAGCTCTCT AGCCACAGTG ATAAAGAT GGTATATATA CTTAAGCCAG GCTAAGGAAA 1380
GCGCTTACCA GAGCTCTGCG TGGAGTACT GGAAGAGAG CTTTATGCG ATCCGAGGAA 1440
CGAGGAGCA TGTAGCGCTG AATTTGCGC ATGTGAGAG AGTCTGTGCG AGGAGAACT 1500
CGAGTGTGAG CAGAAATGGA AGGAGACTA AGTTCTGATG TCAATTTTCT CGAGCGCTTA 1560
CATCTAGCTG TCGCTAAGCG CTGCGCTACT CCGGACTTAA AAGTTTGTG AGCANTAAA 1620
GTCCCTTTCT TGTTTAGAT AATGAAATG AGTGTCTGTT CTGATTAATA TAGGTTATTT 1680
GTATTTCTTT ATTGATTTCT AGAAAGCTT TGTATTTTAA AATCTGAGC TTTATGCTCT 1740
AATTAAGTAA AAAAAATGAG CATGCGCTTC CATG 1774

(2) INFORMATION FOR SEQ ID NO: 95:

316

317

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

5
10
15
20
25
30
35
40
45
50
55
60

GGGACGAGCG AAGGCGAGCG GGCACAGCTC CAGGACTGCA TCTGCTGCG ATTTCCTTTC
CACTGCTGCT TTTCTGAGTC TGACATTAGA AGCCACGCA GAGGAGAT TGAACACACC
AACCTGATTT TCTGCTTCTT CTTTTCATG AATTTCCTG TGGTCTCTGC ACTCTCTTTC
TGTCTCCCGG CAGAGCGGAG TGAAGATGCG GGGCCGACG CTTCTGTGCG GCGACGAGCT
GCTTTTCATG AGCATATGAG TCTCTGTGAT TGTGCTCATC TGGCTGATTT TATAGGCTCT
TCTCTGCGAG GCTGCGACCC TCACTGAGCT GCCCAAGCTG AGATCTGCT TCTATATCTT
CTGCTCTGCG AATGAGAGCA CAGACACCTT ACATGTTCAC CAGTTCCTCTG AGCTGGAGCG
CCTCGGGGCG CTTCTGGTTG GCTTGGGCTT GGCACAGCTT GGCCTGTAGG GCTGCTGCTT
CCTCAGCTTC TTGTCGCGCC AGCTCTCTCT CTTAGCGGAG TGAACATGAG ATGAGAGAGC
GTGCGGCTCG GCAATGAGCT TCTGCTGCTT KTCCTCTCTG CTGCTGCGAG GCGGCTCTCG
GCTCTCTCTC TCTATATGAT GCAATGCTTC AGCTCTCTCC TCCCGGGGCG TGGGTTCTTA
GCTCTGCGCA GCGGCGAGCG CTTACTCATC CTCTTCTCTT TAGCCATGCG TGTGTTCTCT
CTGAGGCTCG AAGGCGCTGA GAGCAAGCTT GAGAGCTGCT AAGGCTTAC GTGATTCGAA
GGTTTCATGT CCACCATGCG TCAAGGCTCG CACATCTGCT CAGCCATCTC ATTATACAGC
TAAAGCTGAT CTCAGCTCC AGCATGCGAA CCACTACAG AGGAGGTGCG GCGCTCTGCT
CAAAAGAGCG CAGGCGGAGC AAGGCGAGC AGGCGACCTG TGACTTCTTA GTACAGATTT
GTCTGTCTCT CAGGACTTCC AAGGCTGCGA AAGACTCTCT AATCATGCGA GCTCATCTC
ACACCAATTC CTGCTTTTAT TATGTGATCT GAGCAATCTT TCTCTAGCT TCAAGAGGCT
GGGAGGCGAG TGAATCTGCT CHTGCGGCGA GACTTCCGCG CTGATTTGCC AATATCGAAA
ATGAACACTA GGAAGACT TACGCGAACA AAGGAGGACA TTAAAGAGC GAGCACTCA
GTCTCTCTCG GCACTGATTT AAGGAGCTTC CACTCAGCCC ACATATGTA GTGGGCGGCG
ATAGCGATTC ACTGAACTC CAAGCCGAGA GCTTCAGGAG TGAATCTCTA GTACTCAAC
AAGAGCGGA CAGTGGGCTT AAGAGACA GCTTTGACTG CTTGAAAGCT TCCCTGGAGC
TGAAGCGAGA CAGGCGGAG GCTTCTGCTG ACATATCTT CCAATGATCA GACCTCGAG
GACTCCAAAT CTTCTCTCTG AAGAGAGCTG GAGGTTTCCG CAGAACAAA CCGTTCGCACT
CTCCACTTCC CAAGCCAGA CTTGGAAGA CATTAGACA ACTTAGGAT TGGGATTTG

1620
1680
1740
1800
1860
1920
1980
2040
2100
2160
2220
2280
2340
2400
2460
2503

TGATATATTT CTAGCACTTG TATATTCGAA AACCTATAG GCAATGATTT ATTCAATAT
TCTGTTCGAA AGCCAGCTG AATACAGAG CAGAGACATG TACTCTGCTG TCAATCTCTG
TCTCTGATTT CTCTCTGCG CTTCTCTGCC TCTTCTTTTA TAGTAGCTG CCGCGGAGC
AAGTACAGCG TGAAGCGAG GTAGCACTTT GCGGAGAGAG GCTCTCTCTG CTTACCACTC
TATACACTGT GGCCTCAACC TCCAGACAG GCGAGAGAC TGTGCGAGC TCGTTTCTTT
TCTAGCTCTG CTGAGAGCT GCGAGCTGAT TGAATGACTC ATGATGAAA CTATTTTCTA
AACAGCTTTC CTCTCTGAG AGATCATAG GCACTTAAC TGTAGCAAT CCACTGAGC
TGGAGTAT CTTTCTCTTT GCAATGACT CTCTCTCTG TTTCACTAG TTGAGCACT
ACATGACAT ATTTTCTGCT GAACTCTGCT ACTGACTAAC ACAGACATTT GCGACTGCG
AGCTTCAGGA CCGAGAGAG GGCACATAC ATACAGAGG AATCAATAG GATCTCATTT
TGAATCTCT CTTCTGACA CTCAGAACG CAGCCGAG GCTTCTGCTG TCCAGAGCTT
CTGCGCTGAG GTGATCTCT CAGGCGCGAG AAGCGAATA GTTAGAGGA CAGAGAGGCG
CAATTCGACA GCGGCGGCG CAGCAGAC CAGTGGAGC TCACTCTCTC TCCAGCTCT
CTGCGAGAC AATTCTATCT ACATTTATG CCACTAGCTCA AAGAGATCT CCGAGATAT
TTACAGACAT CTTCTCTCTC CTCTCTCTCA GCGCTCTCTC TACAGCGAGA GCTGAGAGC
CGCGGCTCT GCGAGAGGCT GAGGCTCGA GTCACTGCT CTC

(2) INFORMATION FOR SEQ ID NO: 96:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2801 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

CTGGAAGCG GAGGCTAGCG GAGCGGGCG GCGCTCTCTG AGCGCGGCT GCTCGGCGTG
CGTTCGCTCT CCGCAGAGCG ACCGAGGAG CAGAGCGGCG CCGCGCGGCG TCTCTCTCTA
TGAAGCGGCA GTGAGCGCGG CCGCTATAG CCGAGCGGCG GCGCTTCTCC CCGCGCTCT
ATCGGAGAG CAGCAGAGCG GCGCTCTCTA CCGAGAGCG GAGGAGAGG AGCATATCTG
ACGTTTCTCA TCGGCGGCGA GGTGCTGCTC GCGGCGGAG CCGCGGCTG GCGCGGAG
GCGCGGCGCT TGAAGCGGCG GATTTCCTCA AAGGAGCGG CCTCTCTCTG CCGCGCGCG
ACGAGCGCG CCACTCCAAA GCGCGGCGCT TCTCTCAGC AAGCGCGCTG CCGCAGCGCA

60
120
180
240
300
360
420
480
540
600
660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
1380
1440
1500
1560

5 GAGGACGCC GCGCGACGG GCGCGATCG AGGTCCCGCG CAGCGCCGAG CCGCGTCCG 480
GCGCGCGCGC GCTCCGAGG CAAACGAGCG CCGTGAGAGC TCCACCTAGT TCACAGATA 540
AATTCGACGA CGAGACTCG GAGTCAGCAA TCGCTAAGCC CAGGTGGTGT GTAGCTCTCG 600
TATTATGTC TAACTGTCT GTGATGCCG CTGAAATTTA CCGTTCAGGT TATCTTCCA 660
GTACAGAGA ATCCATAGAG GATGGTGTG AGGATATACC TACTCTATCA GAATATGTC 720
AGGATTTTT GAATCATCT ACAGACAGC CTGCGAGTTT TGAACTGAA ATTGACAGT 780
TTCGAGAGC CCGTGAATGT TGTGTACAA CAGTGAATCG TTTCAGAA CTTGTGAAAC 840
TCATCTATCA ACGGCGACA TCTATCCCA ATTCTCTTA TATGGAGCT CCGCTGTGTA 900
ATTACTGTC CATTATCTG ACAAATAGC CACGAGTGG CAACTTCCG CAAATGCTAC 960
TTCAGAGTG TCGGACTGAA TATGAGTTA AGNTGAGC TCGAAAGCG GATGAGTTA 1020
CTCGAAAGC ATTTCATGA TTGTACTCT TTCTGGAGA ACTTTATCTT AACCTGAGA 1080
TCAGGAGAC AATGCGAGC GTTACAGAG CAGATATCT TCGGTTGT CTTGCGAAT 1140
TCTGAAATG CCGTTTCTT AATCTATCG ATGACATTT AATTGTGCA GTAAATGT 1200
TAAAGTTGAC AGATCACTT TTGGAGTG CTTCGAGCA AAGGAGAG ATGATATG 1260
AAGAAATAT TCGAGAAAT GAAGAGTTG TCTAGATGC AAATCGAT AGAGTTTAA 1320
AACAGTCT CTTGAGCTT GTAGACTCC GTTCAATTA CTGGCGACA GTCCATGCA 1380
CTTCAGATA TAGAGAGCA ACACAGAA ATGATCTTA CTACTTTATG ATGACACAA 1440
CAATTTATAG ATCTGATGT GTTCTTTCA CTGAGCTGA TCGAGTTAC CAGAGAAAT 1500
ACGAGAAAT ACTTGAAAG GAGGACTTTT TTCGAGTTA TGAGAAAT GGAACGAT 1560
TATCCGCGCG TCGTGATCA TACTTGATG ATATGATGA TGAGATGAC CCGAGATAG 1620
AAGAGCTTA TGAAAGTTT TGTGTGAT CAGAGCTTA GCGAAACAG TAAAGTTAA 1680
TTTCAGATA TCAATTTAT AAGCAGTTT AGTATGTC ATTTAGACA ACACAGAGA 1740
CGAGAAAT GTGTGATC TATACAAAT TAGAGTTT GAGTTATGT ACTATGTAT 1800
GCACTTTAA TTTGTTTAA CACTATGCG CAATATAC TTTATTCCT ATACTTAA 1860
ATGTATAT ATATATATA GTTTATATG TACAGTTAT TCTACTTTT TCGGTGAT 1920
AAATCGAT TTGAATATA TGAAATGTC AAATTTGCG TACTTGTTA GATGCTATC 1980
CTTAAATTC TACTTTCTT GAGGCGAAA AGTCTGTGTC TCGAAATACA TATTACTGA 2040
AAATGTAGC ATCCTTTTT AGTATGAGT ATATAGCTT YCATTTAT TKGACATTTA 2100
GTGTCCCAT GAATGAAT TCAATATGA ATCAATATCT TGAAATCTT TAGCACTTAA 2160
GTCTTGGA TATATACAA ACTGATTTAC ATATGAGT OCTATTTGTA TACGAGGC 2220

TTTTTAATG TCATGGGG GAAAGACCA ACTGTGTGA ACTCCAGCT AAGACACAA 2280
GACTCAGTG GAGATTTAT TCAATCTTA GGAATGTCG TTTTATATT TATTTTTC 2340
AAGTGTCTA CTTCAATAC TTAAAGCTA GAAATTAAT CTGCTTAT TATATGCTT 2400
TCTCACTTT ATTTGTAGC AAGGTGCA TCGACTTTT TACTAGAA TTTTACTAA 2460
TATTTGAT TCAATTTTC ATCTCTTAA TAAATGAT ACCTTGAAG TCACTTTCT 2520
AATACTTTA CTAATATG GTACAGCTC AGCCATATA AATATATT TTAACATAG 2580
TCAATCTTT TTCAAGTAA CTAAAACTG TATACAAAG GATTCCTCT AATATGCT 2640
GTAAATGTT CTTTAATTA CCACTGTTT TCAATTTGT ACATCTGT CTCTAATAC 2700
AGTATGTT CTACTTTC TCGTTTGT TTCACTGTA ATTAATTA GACTTTGAA 2760
ATTAAGCTTA AAAAAAATA AAAAAAATA AAAAACTGA G 2801

(2) INFORMATION FOR SEQ ID NO: 97:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1631 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ATGAGCCAA AGCAATCAC TGATCTTGT GCTCTTACTA TAAATTAAG TGTCTGCTT 60
AATTTTCTC CATACATGT CATGCTTAC AGTATGCTC CATGATGTA ACTGTCTCTC 120
GAGCTGCTTC TCGTTGAGT TGTCTTCCA GCAATTAAGC AAGAGAGCA CAGAGGCG 180
TGGCTGAAGG GCGTGTGCG AGCTGTGACT GTACCGGCG GATTACTTCT GCACTTCTAT 240
TCTTATTTAT TCGAGACCA GGAAGAAAT GAAATCATG CAATCTACA AATTATCAAT 300
AATCAGCATG CTGAAATTA CAGGCTATT CCGTGTGTG GAGAGGCTT TCAATGAGCC 360
CAGCAGCCA TACTTCAGCA GCGAGGCTT GTTGGTTTC AGCTTTACG CCGAGCTTTA 420
AATTTTCAG TGAAGTAT TCTGTGAT GTCTTCAAT GTATATCAAT ACTGATTCG 480
AGGCTCATCT GCGTACTTT ACCGATTTT CCGTGGCTT GATTATGTC GTTTTGAGG 540
GCGAGTCCA AATTCATGA GCTCTAGCA CCGTGTGTG GTCTCTATGT TTGCTGCTA 600
ACCAATGAG CTTGTAGCTT GATGTGCA TGTATGCTC AGGAGGCG AGTATCTTC 660
CAGAGGTTA AAGATGTC TCTATGAT ATGAGACTT TAAATATG CCGTCTGTG 720
GCTGAGTTG TCGCTCTCTT TCTGGGCTC CTTTGTAGC TGTCAATGT CCGTCCCTG 780
AGGTTGCTT TGAATCAG TCTCTTTT TATCATGAG AGGAGTGGC ACTTGAGTC 840

5 CTGATGCCA AATCATCTC AGCTATACCA TTGATGGGTC CTCAGTGGTG GTTGAAACT 900
GTATTGAC AGTTTACG AATGGGATC CGGACATG ACCTTACTA TATTTCTGT 960
AACTGGCAG CTCGGCTGAT CTGTGGTG TTGCTTTCC TGTGTGAC TTATGTGTA 1020
GGTTCTGGT TTGTCTCTT ACTAGTGT ACTGGGAA TGCAAACTT AGTCCTGG 1080
CGGATTTATC CAATTACT CATGGTGGT GTATTGATG CAATTGTGCT GTTCAGATC 1140
CGGCTTTTA AGCGCTTTA TGAACATAT AAAATGACA AGTACTTGT GGTTCAGCA 1200
CTGTGAACT AGGAGGGA ATCTGGGA CAGGGCTAT CTCACGACC TGCAGATCA 1260
TCCAGGAT AAGATGTTG TCTCAGAC TTGACTTCC CCTTTACATG TCTTTTGTG 1320
TGGACTCTC TCTTTGAGA TTATGGCCAG TGAATCTCA GCGTTGTATT TAAATTAAT 1380
GTATTGACT TGTGTCTCA GATTCAGAG AGCAGCGTG TAAGATTTCTG CTGTCTCC 1440
TGGATCTCT GACATTACT CTGTCTGAGA TTGTGTATG TGTAAATACA AGTTCTTTGA 1500
TACCTTAAA CTTTGATTA AACAGATGT GCATTTGACA TCTTTTAAACA AATATATAT 1560
TATTTTATA ATCTATGTTG TCACTTTTAA AAAAAAAAAA AAGGGGGGCC 1620
GTTACCGAA T 1631

(2) INFORMATION FOR SEQ ID NO: 98:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

40 CGAGCTGG CGAAGATGAG GGGAGGAC GAGCGGCC GGTGGGGTT GCTATGCTT 60
CGCAGACT ACTGAGCAG CCAGCTGAGA AGATTTGAG GAAATGCTG CTCCTGGTC 120
TCCAGCCGG ATGATTAAG TCCAGCCGA AATTAACAT CGCCCTTCT GCTTCAATG 180
GAAAGCCAC GTGAGATGC TGGCTGGA TATATACAC TCACTGTAA CACAGATAT 240
CATCTATC GTATCTGTG TGGCACTAT ACCAGAAC ACACATTA CAGTTGGTG 300
AGCGGTGTT GCACTTGA CAGCATATG CTGTCTGG CAGCGGCC TTATTTACG 360
GAGCTCTG TTCAATCCA CGGCTCTTA CAGAAAGAG CTGTGCAATG AAAAAAAAAA 420
AGTTTGTAA TTTTATTA CTTTATATG TATATATG TATTAACAT ATTTCTGTAT 480
TCTTCCAAA AAAAAAAAAA AAA 504

5 (2) INFORMATION FOR SEQ ID NO: 99:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

10 GGACAGGG AGGAGGCTT CTCGTTGG TACTCTTGT GTGCCCTTA GACAGGCTG 60
CTGCGCGTT CCACAGGTA CATTTAGGAC TTGATCTTT CTTTTCCTT TTGAGTTGG 120
TCACTGATG ATAGGATAC ATGGCCCTT AGGATGACC CTTCGAACTG TGGCAGTTC 180
CTTAATCTC AGCTGGATC CTGACCTTG GAGGCCCTG TAGGGCCAG CTCGTGAAA 240
ACCTGGAT TCAATGCCA GCTGTGGAG AACTGTGTC GAGGCGAGG TGGCTCGAA 300
CACTGATGT TCTGGGCTG GAGGAGAG GGGCTCGGC TTTCCTGAA ATGAACTAG 360
CTTTTCACA GTTCAATGAC TTGTCTCAA AACTTTTCT AATGATTCG TAGGTTTTCA 420
TAAGATTTT TCTTTAAG CATGGAAG GAGATGCT CAGGAACTC ATGTTTGTT 480
TCACTGGAT GGGCGCGGT TCTACTCT CCGGCTTCC CTTCATGTC GCACCTTGT 540
CGAGGCGCA CGAGGAGAC TCTTCCACC TTCTCCACT GAGGACCA GGGGCTTGA 600
ACCTAATTT GCTAATCAG AGGCAATTT TTGTCTGAG TATCTTTTAC ACTGTCCAA 660
CGCTCTAT TTTTAAAG TTCTGTGCT TGTATTACA CAACTTACA GAGAAATAT 720
TTCTGAAGC AGTTTATGT GAGATTCGC AAGGGAGGT TCGGTAGCA AATATATTA 780
GCTGTTTAT AACTGACA GGCAGACAG CAGGAGCA TTGAGAGCA ATTTGCGAA 840
GATCTACCT GAGATAGCG CTGTCACTG TCTTCAAC CTAATTAAC AGCGTCCAA 900
AGTGTTTTC TCTTTTAAA AAAAAATTC CACAGCTTT TAAAGTACA TTTAAGATC 960
CATGTACTT TAGAATGAA CTCGCGGCC TGGCACTGT CAGGTGCT AGAAGTTGG 1020
ATGCTCTGG AATCATGTC ATACTCATC CAATTGTGT TCTTTAATG CATTTTGTT 1080
CTTTTACAG ATCTGCTCT GTGGTGTG TCTAAGAGT CGGACACTT GGTTTTGTG 1140
TTAGATTGAG CTGGCACTT GAAATGACT TCTTTATATG CAATTTAGC AGGACCATC 1200
TGTGTCTCT GATTGCTGC TATGATGAG AGGGAGGGA GGGATGTGAC CCAAAAGTA 1260
GGGCTTCCA TTGGTTTGG CAGGCCGAG CACTTCAAT GATTTACATG GTTCTGTGA 1320
AATTTAAGT TTAATGTAT AAGCGAGC TGTTCGTG AACTGTATA TTTTGTAAAT 1380
AATATATG CTACTTAAA AAAAAAAAAA AAAA 1416

5 (2) INFORMATION FOR SEQ ID NO: 100:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2847 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

15 GGTATGAGCA ATTGTGTGC TTATCTATC TCTCCAGGA CTGGAGATT TGGATACCA 60
TTATTTACA CACCAATCA TATCCAGCA AAGTACCCTA AAGAGAGGAC CAGTGGCAC 120
TCTCGAANA ATTATAGTAT CAGAGATTA AAGAGATTT AGATTTTGA AGCTTTATT 180
GTCTTTCCC AATATCAAT GTTTGATCT CAAATAGTAG CTTTATATA GCATATGACA 240
GATCATGGT TCTCCATATC TGTATCATG TTACTACTTT GATATCAGTA TTTTGGGAAA 300
TTGAGGANT TATCGAGTGG ATATTAATGG AATATTAATA ATATTTGCCA ACTGTCTCA 360
GTACTTATC ATATCTCTGT GATCTCTCAA GGAAGGAGT TTTGCTTTTA CTTAGAGAGC 420
GTTTGAGATT TCGTTTATAG ACTCTGCTG TCTTCAATAC CTGATTAAGC TTTTAACGAG 480
GAGCAATTA ACACAGTGA GCACCTTTTG CCGAGGCTC TAAATGCTCG CCGGAGCAT 540
TTATCAATGT AAGAACTAG ATGCTTCTG CAGTGGAGT ACCTTCCCT AGAGCTGGAG 600
CATCTGCTT GCGCTTAGC CCGAGCATGA TGAAGCTTCC CTCCTGGCAG GTCAATTAHA 660
GTTAGAGAG TCAGATTTGG GTCTTGGCTG GGTCCAGGTA GAGGGTTTG CTGAAACCCC 720
TAAAGAGAG TCACCAAGG AGGCAGGTAA TGAATGTTT CAGATCTGAT CTGATATCA 780
TAGCAATTC TGGATATCT TCAATATTTG AATTTCTGTA TCGTGAAGAG GACTTTGATT 840
TGATATCAT AATATCAGA CAGTCCGAG AAGTGGCTTG AGTCTGGGCT CTGACAGCCC 900
AGAAGGGAA ATATCTGTA TTAAGAGCA ACTATGAGC AGGCGCTGAG CTGTCTCTTA 960
GATATTAHA CAGATGGGA GTGAGAGAT CAATTGCTC AAGTTATACA GCTAGGAAT 1020
ACTAGGACA ATCTGAGC GCACGCTCCC CTAAATCTGT GAGAGGAC TTTGTACCAC 1080
ACACCATGT CCACCTAHA ACAGAGAGT AAGAGACTT CAGGTTTTCC CACTGTGTGC 1140
TGACATCCC AATTAATGA TCTTCTTCA AATGAGTTT CAGATTTATA GTTAGGGCTC 1200
AGAAATGCA TTGAGTATG CTTATTTTC CCGTTTACA GATGCTTTAA GTACAGATTG 1260
CTGACTTGA CCGACCCCA GGAATGAGA GAACATTTCC TTTTCAATC CATCTTCAT 1320
AATTAAGGTT TTTCTGGC TTCAGAGTA TAGACTTTG CAGCAGTAT AAGATGAG 1380

5 GGTGTTCTGC TGTCTACTCA ACTTATTTG AANATGCTG CAGCTTCACT CCGTAGAAA 1440
AGGAATCTT CATATTTTAG TAACTTAGC CCGCAATGTA CTCTGTAGG ATGTGCAAT 1500
TAAAGTCCA GTGAATCTGG CTCTTACTT GATTCCTGAT TTTATGTGAT GTTCTGGGG 1560
ACTGTGTACC TATATTAHA GGAAGATGT CATATGTGT TATATGTATA TACATACATA 1620
CAATGTGACA CACACACCA CATATTTGA GAGCTAAGA AATCTAAG CAGCCCTTC 1680
ATTATCTTGC GTACTACTTC AAGATTTCT GTGAGGCTA ATTACAGTG TCACATATA 1740
GTTGGGCTT AGGTACTTGC TTACAGGAG AGCATTTCC TACGAAGGT CATTAAGTCC 1800
TAAGGAGCT AUTCAAAATG ACGGCTCTGA AGGAATTTGC ACTCCAGCC TCTTCAGGA 1860
TGTCTATATA GATGGGAAC TTGGATGCC AGCATTTTG GTGACTGTAG AATCTACTA 1920
CTCAGTTAG ACTAAGGCG ACAATGCGA AATCATGAC CTTGTAGTGT TGGAGGCTC 1980
TAGAGCTCC TCTCTGCCA AGTAGAAT ATCTCTTGC CATCTCTGA ATTCCAGATT 2040
CATTAATGCT GTGTGCATA CATCTCTC AATAGAAA TTAAGTCAAT GTTTACTGTG 2100
TGTGTATCA ATCAGATTTT TATGTTTAA AATTTCTCAT TATGATTTGA GTCCAGGCA 2160
GCTTAAGAG AAAAGAGAG CCGATATGG AGACTCAAT CTCATTTATA TTAGCTTTAT 2220
CGAGCTGCT TTTTAAGCC CCGTAAGCTG TCGATTTCA GAAACATTA GACTGAGGCA 2280
GTTCTGATT CTGAGGCTT GCGTGTAG ATAGATAGT ATAAATTTGA ACTGAGACA 2340
TACGAAGC ACAGAGCA GGGCGAGCG AAAAAATGA AATTAATGAG AGACATTTAT 2400
GATTAATGTT GTGAGGCTT GCGTGTAG ATAGATAGT ATAAATTTGA ACTGAGACA 2460
GACAGAGAT TAACTTTTA TGAAGGTTA TGAAGGCTG CCGTATGAT ACTTTAGGG 2520
ACATGCTAG AATGGGAGA ACGATGAT CCAATTTCTT CCGCTTAT TCTGTGTGAC 2580
AGTCAATTC CCGCTTAGG AATCTTCCC TTCCAGCTT TACATTAAC AAGGAGAGC 2640
TGAATCTTC AAGGAATTA CAGTTTGGG TTAATGTTT AGTATATCAT TTTCAATGCT 2700
TAAATTTAT TGTAGAGAG ATTACTGCT ATCCAGGAT GTTGGAGCTT GGTGCGCTG 2760
TGAATTTGA ATCATTAHA CTATTAAGT AATGCGAA AAAAAAAA AAAAAAAT 2820
NAAAAATC GAGGCGGCC CTTAGCC 2847

(2) INFORMATION FOR SEQ ID NO: 101:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

5 GAGTGTGAG GAGGAGGATA AATATCTAG AGGAGAGAT TCACTAGAG CGAGAGGCA 60
CCCCGAGAA AGGATATGAG GCTATCTAT CTTCTTTACA AGGGGTGACC ATGACTTACT 120
GTTCAGAGAT ACTGAGATGTA TATTATATGT TCAATGTGCA ATTATAGTGA CGAGAGGAAA 180
10 GAGCAATTTT ACTCTGTCC TTATTCTAGT TCGTAGAAG CTGTGAGACA AATATATAG 240
AATAGAGAG GCGCTTTCTT TTGTATATTC TCGTTTCAAT GCAATATATTT TTATTATACC 300
ATTATTTTCA AGAGATTTAG CATTGCCCTC TCGTACTTTT TTCTTTTATC GAGCCCTTTC 360
CTTTTTTTGA AGGGGGGTGA TATATAGAG TTCAATGAG AGGTGCACTG AGGCTGAGAT 420
AAGAGGAGCA GATAGAGGAG TTATCTTAAAG AGCACTTAT TTCTTTGAG GCTTTTCTAG 480
20 AAGAGATAG GGTTCGAGAT GGTTTGATAT TCGCATGATG TTGAGAGGCA CTATATGTGG 540
TTGAGATAG AATATATATTT TCGCATGTGG GAGAGAGAA TTCTCTTTAG AGGGTGGCAA 600
AATGCTTTTG CCGAGTGTCC CTATTTTATAG GATCTTTTCC TTCTTATTC CTTCGAGTCA 660
25 GGTGTGTGCC TATACAAAG TTCCGATGAG TTCTCTGCA TATTCGCCAT TTGTAAATGA 720
TCACTTCTCT TTCTTAAAC CTTTTCCTGT TCAATGTAT AGAGATTTG GAGGATGAG 780
30 ATCTATGAT GAAATGTCCC TTGTTCATCT GTGTCTCTTG GAACTATGC TCAATGAGAA 840
TTCTGGGTAG GAGAGAGGTA GCTGAGATTT GGTGTGTGGA CTGAGATAG GGCATTAACC 900
ATCTATGATA TTCTAGCTCC CTTCGAGCA GCGCATCTT GCTAGTGTGC ACAGAGGAAA 960
35 GCTAGCTGCC TTGTGGGCTC AATTTGCCCT CCGCTGATG AATGTAAAG AATATATATTT 1020
TTTCTGTGAG GTCTAGATTT GCTGAGACA AAGTATATCT AATATATGTT TTATGTGGTGA 1080
40 TTGTGTGAAA ACTGAGAGAG CTCACTGCCCT ATTAAGAGAA ATTAAGAGAG AAGTGGAGGA 1140
GAGGAGAGAA AGAGATATTT AATTGTATGA GATTCAGGCA TCCCAACTCT TCTCTCTGCA 1200
GTCCCTGTCT CTGATTTTTC TTGTTTGTAG AATGCTTTGT GGCAGAGACC AATATATCTTT 1260
GCAATCTGAG AATCTGTGAA GAGGGATATG GATTCAGCA GTTGTGTGCA TTGTTTTATT 1320
GCAATCTTTC TTAAATATTA AAAAAAATA AATTTATAG TTATATCTTA AAAAAAATA 1380
50 AAAAAAATA AACC 1394

(2) INFORMATION FOR SEQ ID NO: 102:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 794 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

5 GAGAGAGAG GAGATATAGAG GACTATAGAG AGGAGTGTCC GCAATATATC TTATTTCCCT 60
CGCTCTGTCC GCGCCGATAT CTCTTTTACG CCGTTGTCCA CCGTGTGTGG GTATCATGAG 120
10 CGAGAGGTAG GCGGCACTGC AGTTCATATC GATCTGTGAG TGTGTCTTGG GAGCGAGACC 180
GTCCGAGGCC GGGGGCGAGG GAGGCCAGAG AGCTGTGCC GCGCTGAGGGA GAAAGAGCTG 240
GAGCTGTGCC TTGTGGGTGAG AGGATATGAT TTGTGTGAAA GATGTGAGAA AGACTGTGGTT 300
15 TTCTTTTGA GAGGTGATAT ATGCTGTCTT CCGTGCAGAG TTGATATGTC ATCAATGTGG 360
GTTCCTTATC TCAATCTGTGC TCTTCTCTCT GTACAGATCA TCTTCAAGAT CTATAGATGCC 420
GTATTCAGAG CTGTACAGAA TCGAGAGAAA GCGCATTCAT TCGAATGACT ACTGTGAGAT 480
20 AGAGATATCT CTGTCTCTAG AGACTTTTCA TATATATATG AATGTGTGCA TGTGTGAGAT 540
GAGAGGAGCC CTGAAAGTCA TTATGTGTCT CTTTCTGTGTA GAGATGTGAG TTGAGTCTCT 600
25 GAGAGTGTCT GTTTTCAATGT GGTATATATC CTATGTGTGT GGTGTTTTGA AGGAGATGAC 660
CGTTTATAT CTGTGTAGAG TGTCTATTTT GATGTGTCCC AATGTGTATG AGAATATGAA 720
30 GAGCGAGAT GATGTGTATG TTGCAATGCC CCGAGATGAG AGCAATGACA TTGTGTAAAA 780
GATCCAGAG AAAA 794

(2) INFORMATION FOR SEQ ID NO: 103:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1544 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

45 TTGTCTGTCT AGTGTAGACC AAGAGATGT TTGGGCATTT GGTGTGTGAG CCAATTTTGG 60
AGCAAGAGAG TTGTCTTCTT CATTGCCGCA GGCAGAGAGC TTGTCTGTGG GCGAGCTTTC 120
50 CTGAGTGAAG AGAATATATC CTTTTCTTGG GCGCTTGAAG TTGCAATATG AGGCATATAG 180
GCGAGAGCTG CCGTGCATGC TTGGGCTCAG CCGAGATCTG GATGTGTGCTG GATGTGTCTC 240
TTTGAGATTT AGGCATCTCC TTGTAGTATG GATGTATAT GCGAGCTTTC AGGTGTCTTC 300
55 AGGTATGAGC ATGTGTGGGG TCAAGGTATCA CTGAGCTTTC CTGCCAAGCT CTAGAGAGCTT 360
TTAAAGATAG TTCTTTTGGC GCGCCCTTGA GGCAGAGAGC TGAAGTGAAG CTTGAGTGAAG 420
60 CTGCCAGAG GAGAGAGAG CTGAGCTCAG GATGTGAGCA CCGAGACTTG GCGAGGCTTT 480

5 CAGAGAGCCCT GTGTGGGGCC CCCAGGATC CTTAGCTGAA GCGGGAGAG TCATCTCTCA 540
TCTCAGGAAA TTCTAGCCCTT TCCCTCTAGG GAGCCACGGT TGAAGGTTGAG CCCCAACCC 600
TCCCTTAGGG CCGTGGGTTGG GCAAGTCTGG CCGCTTGGGGT AGGAGAGGGAG ACTCAGGGCC 660
ACACTTGGGT ATTCTTCAAT TTGAGAGAAA CACACACTCA CCGCGCACTC ACTGATTCCT 720
ACACATTTGCC AAGATTTCAC ACATGTGACC AGGGGCGACC AAGTTCCTCTG TGAACCTTTGT 780
GACTAGATTC CTATTTTCTC TATTTTCTCC TGGGTGCTGG GTTCTGTGTG ACTGCGGGA 840
GTGTGATAA TTTTATTTTC TGTGACATG TTTTATTTGGG GTGGCACTG GTTCTCCGAT 900
CGCTGGGCTG GTGTGAGCC CAGGACTGTA GTGCTGGGAG CAGTAAGCT CAGCTCTGTG 960
TAATGATGTA TGTATTTGCT TGTCTGTGTC TTATGATCCA ATCTTTTCTT ACATGAGCC 1020
TGTGTTTGT TTATGCTAG TTTTATCTGG CTTGCTTATTT TCTTTCGGG GAGGAGGG 1080
TTTCTAATC TGTCTCCAG CCCACTTAT ACACACCCAC CTGCTGGGA CTTACTCTC 1140
GGAGGCGAC AGCAGGGAG CCGCAGAG TGGGTTCTG CCGCTGTGCT GGGGGTGGG 1200
GGAACTGGG GGCACATGTS GCGCTTGGCT TCTGAGAGC TCTCATGTCC AGGCGTTTGA 1260
GACTTTTCCA CATGATAAA GAAAGGGAG GTACAGAGT TCCATTTCC TTTTATTTT 1320
GCTGGTTGCT ATCTGTAAAT GTTTAATAA TATCTGACA TGTATCTATC AAGCGCAAG 1380
ATTTCUAGT CTCCTTCAC ATATGTGGC TTTTACGATG TTTATTTCC TTTATTTTAA 1440
TTTATTTTCCA GTTTTTCAG GGAATAAG TCTGGAATTA TGAATACAG TTTATTTTAA 1500
ATTGTTCTT GATATAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1544

(2) INFORMATION FOR SEQ ID NO: 104:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 871 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

50 ACCACGGCT CGGCTGTGTC CACCCGAGG GTTGGAGTG AGTACCAGA TTCAGCCAT 60
TTGGGCCCCGA CGCTCTGTGT TCGGATATCC GGGTCTGGG GATTGAGTTC CCGGTTCTTA 120
AGTGGGCTGG CTGTCCACCC GGGGGCTGG GAGTAGGTA CCGATTTAG CCGATTTGGC 180
CCCGAGGCT CTGTCTGG AATCCGGGTS CTGGGATTTG AGGTCCGGT TCTTACGGA 240
CTCGAGATG GAGGAGGCG GAACTTAGG AGCGCTGAT AGATGCTCC ATCTACTGT 300

5 CTGTGAGGT GCTTGGGGA TGCATATGTG GTTGAAGTTC GTTCTAGGCT TTCTGCTTTT 360
TCCAGAGCT TCCCGAAT ACTTTCGAC TAGTCCAGAG CAAGCTCTTC CCGTTCTACT 420
TCCACATCTG CATGGCTGT GCTTCTATCA ACTCTGCAAT CTTGGCTTCA CAGATTCCTT 480
GGCTGAGCT CAGATTTCTG GAGGCGAGCC AGCTTTTACT GCTGTCTGTS AGCCTTAGGC 540
TGGCACTGT CAGCGCTGC TGGCTGAC CCGCGACAC AGCTGCCATG TGGCGCTGTC 600
AAGCCTTGG AAGAGAGG AGGCTTGGT GGGAGGTAC CAGGAGCCA ACAGTTTCCC 660
GATCTTAC GCGATTTGG AGGAGAGAC CCGCATGTCA GTGCTCTGCG CCGATTTTTC 720
TTCTGCTAC ATGGGCTGTC GTCTCTTTC ATCTGGGCT GGTTCCTTAG CAATGGCTC 780
TGTCTGGTG GCTTGGCT GGAATTAGG AGCCTTAGC ATGGGCTG CAATGATTA 840
AATCTTCTT CAGAAAAA AAAAAAAAAA A 871

(2) INFORMATION FOR SEQ ID NO: 105:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

35 GCGACAGTT ATAGCATGCC ATTCTACTT TGTGTTTAT GTCTCATGAC TTTTGTGAT 60
TTAGAACAAA ACAGTGCAC CTTAGAGCTT TGTGCCATG AATTTTCCA TCTCTCCA 120
AACTGCTTG AGTTACTAG AACTTCACC TCCCATGCA CTGAAGGCAT TCTTTTCA 180
AGATACCAGA ATGGGTACA CATTTACTT GCGAAGCTT GAAGAACTCT TAATGTTTC 240
TTTTTATTA GATGAGGCC CCACTTTGGG GACTAATTT GTGCTATGCG CCGAAGGAG 300
TCTAATATTT ATTTTATTA AAGAGAAC TCCCATTA TTTTGTGGG GTTGTGTTT 360
AATTTTAT ATGAAATTT TTTTGGGCT TTTTGGGCC ATGG 404

(2) INFORMATION FOR SEQ ID NO: 106:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1542 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

60.

5 GTGAGACAG TGGAGCCGC GGGGAGGAG TCTGAAAGAG CGAGCTTCA GAGAGGAG 60
GGCTCTTCA GAGGAGAG GAGAGCCCTG GAGAGGAG GGCTGAGGT CCTTAGCCAG 120
GATGAGGCT GTTGTACT TGTAAAGAG GTTGTAGAG CAGGAGAT CCGGATCCA 180
GGCTACCT CTGATGGGT CCCCTTCT AATGACTTC ATTCTCTCA CTAGCTGTA 240
CTTCCTTTC TCACTGGGC CTGCTGAT GGTATATG AAGCCCTTC AGCTCCCTGG 300
CTTCACTAT GTCTACACT TCTGCTGCT GGCATCTTC CTCTACATG TCTATGAGTT 360
CTGATATG GGTCTGCTA GACATATAC CTGGGCTGT GACCTCTG ACTATTCGA 420
CAGCCCTGAG GCACTTAGCA TGTTCGGGT GGCCTGGCT TTCTCTTCT CCAGTTTAT 480
TGAGCTATG GACATGTA TCTTATTT CTGAAGAA GAGGGGAG TGACTTTCT 540
ACATCTTTC CATCACTG TCTTCTCTG GAGCTGCTG TGGGGGTAA AGATTCCTC 600
GGGAGATG GCTCTTTC ATGCAATAT AAATCTTTC GTGATCTA TATATTAAT 660
GTACTAGGA TTATCTGCT TTGGGCTGT GGCAGACC TACTTTGCT GGAAGAGCA 720
CATGAGCC ATTCATCTA TCAATTTT CTGTCTCTA CTGCAATCT CCAATTA 780
CTTATCTTC AGCTTAAT ACAGTACC AGTCAATAT CACTCATCT GATATATG 840
CAGCATCT TTCACTCT TCTCAAT CTGCTATCA TCTTATCA AGGGAGAG 900
CTCTCCCT GCACTTAG AAATGAGC TCCAGTAT GCAAGCTCA AGGCCACTG 960
AGAGATG CTATATAG GCGCACTA AGTCTCTAG GATCTACT TAGGGCTG 1020
TCTCTCATG CCTCTCAC CTAGCTAT GACCAAGCT TATGTCTCA GCACTAGCA 1080
GGGACTGG CCTCCCTCC CAGAGTGC TCTAGAGGA CAGGGCTT GTTCTCTAC 1140
CACTTCCC GGGGAGCT CAGGATAT GCTCTATTC TCTCTCCAC TCGAGACTG 1200
GGGCTAAA GGGCTTACA GTTATTTCC CTCTCTGCT TTAAACTTG GAGAGGAGC 1260
ACTAGGGCT GGGCCAGAA AGGCTCTGT GGCCTTTTC CTGACAGA AGAGCTAGC 1320
AATATCTCA CTGTGACC AGTCTACT CTCCAGCCA CAGCTAG CAGTACTTC 1380
TGGCCAAAG GTAGGCTG GGGGGGCT GGGATAG CTCTGGAG CTGCTACTC 1440
AACTGTCT TTATTTAA GTGAGAG AAACAAA AAAAAAAAA AAATCTCTA 1500
GGGGGGCCG TACCAATC GCGGTATGA TGTATGAA TC 1542

(2) INFORMATION FOR SEQ ID NO: 107:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2127 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

5 GGTACTCAN TCCAGTAAA TAGTCTTACT GGAACAAG CCTTTATCA AGATATATA 60
ACTCTTCTT TTTCTTTTG GAGAGTCTT TTTTCTGTA TGGACCAAT TCACTGAGC 120
AAGCAACACA GTATCTTAG CAGAGATCG GCACTTAGG CCAATTTGG GAGGCGAGT 180
RACATATCT GCACTCTGA GTTGGAGAA TATGACTCC ACTCTTCACT ATATTCAGR 240
CAATTCAGC TTGACCAACA ATAGCAATTT TAGCCCTGAT GAGCAAGCA GACTAAGT 300
ACAGATGTT GTACTTAGG GTTGTAGA TCAATATTA TCTATGACT ACCCTTCTG 360
TCCAGAGCG GTTGACACT AAATCTTAA GCACTTGA TATAGCTTC CTGCTGTC 420
CTTGACACT GGAAGAGCA ATGCACTG CCTGAGAGAG ACATATGGA CTATGGCTC 480
AGCATGCA TCGAAATTC GAGCAACTC TAGCATCTC CATCAGGCG CTTCAGTTA 540
TTCTTGAGA TCAATGACA GCTGAGATC TGTTCATAT TTTTATGTA TTTTAAAG 600
ACTTCATGA AGTCAGATA GTTCTCTTA AACACTGCA TGAATTTCTG AGCTTCTC 660
ATATGAGAA AGAGAGAAA TATCTTATC AACTTAGGA GTTTTGTG ACAGATATA 720
GTAGAAATG CCGTTTGA GCTGACTCG CTGAACACT GATTTTACTT CTAGATTA 780
ATATGTCAG AGATTTTAT GACTTTTAC GTCCATTC TCTGATCTG TGTGAGACA 840
AAATTTCTC TTTTCTTGG ATTTCTTCA AGTGTCTAG CAGATGCTG AAGAGCTCC 900
AGCGGAGAC ACCAGACAG TTGAGATGG ACCCTATCA TGAATCTTG GAGACTTTG 960
GAGATCTCC CAATGTCTT GTTGGGAG CCTTTTCTT TGTCTGCCAG ACTCTCATG 1020
AGCATGACT CCTTCCATG GACAGTTTG CTGTGCACT CATCCGCA CTCTTACT 1080
TAGCAATGA CAGGTCTCT AAGTGGAG TGTCTTTC AAGACATTA AGACAACTC 1140
TACTAGAAA AGACTATTTT TGGCTCTG CCACTTCCA CAGAGGCT GTGAGAGGA 1200
CAATCTGCT TCTTAGATG GACCTTACA CCAATTCGA GTATTTTCA AGCATCCAC 1260
CTGCAATAC CAATATCTC GAGATGCA TGAAGAGAG GTTCTTACC TACTAGAGG 1320
CTTGAACTC GTTGTCTTC CTGCTTCA CTAGAGGAG GTTCACTGG CATTCGAC 1380
CAATGTGACC TGGATAGCT TTGGGGGAG GAGAGACTT CCTCTCTTC GACTTCAAT 1440
GAGGTGCA GTTGTCTCA CCAATTCCA GGAATTTCA GAGTCAGAG AAGTACAT 1500
AAAGACTAT ATCTTATCT GACTTAAAG KRNAAKMMW KCTAGNSRA TTTATTTSM 1560
CMMRGASH WYAAWSTK SAGCTTACC KSRSTGRKIG HBCCTCTAGA ATTTGRLAK 1620
CMTTUSCT RMGGAUKS GGLSAGCC AGAGACTGC ATTCCTTCT CCTGTTTTA 1680

TTTAAACAT GAAATATGAA ATTCTTACAG CCGTAGGCGA GAGCTGTGCG CAGATGTGAA 1740
AGAGACTTC AGTATCAGCC CTAACTCTTC TCTCCAGGAA AGGACTCTCT GGCCTCTGTG 1800
GCGAGCTGTC CAGCCGAGCC CTGTGTGTGA ATCTGTTTGT AGCTGTGCGA ATCGGAAAGG 1860
AGGGGTCTTT ACATCTCTTA AAGGACTCTA TCGCAGCACA AGTAGAGTGT ACTTAACATC 1920
TTAAGCCGAG CATATCTGTC TACACATTTA CAGATCTGTT CCGTAGGTGTC TGTGTCTGAT 1980
TTTTCTATCC TGTGTATGAC CTGAGCGAAA TTATTTAGAC GTATATATGA TGTCTGTGT 2040
TTTTAACTTG ATCATGATCA GCTCTGAGGT GCAACTCTTT CACATCACTGT ACATACCTGT 2100
GACCACTCTT GCGAGTGTG CAGTCTTTAA TCATGCTGTT TAACTGTGTC TGGCAGAGT 2160
TCTCTTGTCC AAATAAATTT TATTATTAAG ATCTATAGAG AGAGATATAT ACACTTTTCA 2220
TTGTTTCTTA GATGCTACCC ATATATATCA ATTGTGAGCC TGTAAATAAA AAAAAAATA 2280
ACTCGAGGGG GCGCCGGTAC CCAATGCGCC GATATGATCT AATCATC 2327

25

(2) INFORMATION FOR SEQ ID NO: 108:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1062 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

GCGCCGCGAG GCGCAGAGC GGTTCCTGCA GCTCTGGGTC CAACCGGACT AGCAATATC 60
TTCCGATCCC TCATCATGCT CTTCCTGATC CGGATCTCGG TCCAGGTGCC TCTGCGCCGC 120
ACACAGAGG TCGCGAGGT CCGAGGTGTAG TTCTCTGCGA GGTTCCTGGA GATGCTCTTC 180
CTCTTCTTCC TCATCATGCT CTCTCTGCTC TTCTCTATCC TCATCATCCA GTTCTCGAG 240
CCGCTCAGCA ATCCGATCCC CCGCCGCGGA GAAATGTACA GAGAGCGCGG GTACAGCTCT 300
TATGCTTAC ATGACATTA CCAAGGCGCA AGATGCTTAC AAGAGGAGG TCGATATGAA 360
GAAGAGGG TGGTCTTCA TCGAAGATA CCGTGGCGCA TGACTGTGATC AGAGCTGAAA 420
CAGAGTCTT CCGTTTTCG AGAGTATGAG GAGTGCACCA TCCACTTCCG TGTTCGAGGG 480
GAAATACG GCTTGTGAT TTATGCTAT GCTGAGGAGG CATTTCGAC CATTTAGAT 540
GGGCGAGCG TCGCGGCGCG AGATGAGAG CCGTTTGATC TCTGCTTTGG GCGCGAGGG 600
SMTNCTGCA AGAGAGCTA TTCTGATCTT GACTCGAACC GCGAAGACTT TGACCCAGCA 660
CCGTAAAGA GCAATTTCA TTCTCTTAC TTGTGACAT TGTGTAAACA GCGCGAGAG 720

60

ACCTCAGGA GGTAACTTC GGCCTTCC TGTATGCTT TTCTCTCTTT GAGGTGTCCC 780
AACCTCTCC ACCGCTTCC CTTACTTAG GCGAGAGAGC TGTACTTAG ATACTGTCTT 840
TATTAAGAAA TCGAATAGAA AATATGTTGA ATCAGATTTT TTAAGAGGGG 900
TATTTGTTTT TTATTAACAG GTATGTAAAC AAGTTAACTT CCAATCTCTAT GTAAATAGG 960
AGGGCTGAG GCGATCCCA GTGTTTGGA CATAGCTCAC TATCGAGACT AATTAACATC 1020
AACTAGAGAG NAAAAAATA AAAAAAATA ATTTAAAAA CT 1062

15

(2) INFORMATION FOR SEQ ID NO: 109:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

GAGAGCTCA CACTCTTTT CCAATATCAC TCAGATGTA GTGACATAG CAGCGGAGAA 60
GCACTTACT GTTGTGTGA GTTGTGTGA TCAATCTCAT AACCTAAGAG AGGATTTTAT 120
AGGCTCTCG CTTATGTAG CCGATCGAGA AATTGTGCT GTGAATTTT ACATATATAT 180
AACTGCAAG TGGGATTA ATATGACTA TTGTGTGCG CAGGCTTACA TTGCTCTAG 240
TGATTTTCT TCMAATGTA AAGTTGTGC TTCTAGACTT TTAAGAGGGA TAACTCCGAA 300
TAACTGAA ACATCTGSM CTTCGCTGCT SKTRRRKRYC GTATATATTC WARRKSAKYM 360
CCGCTGTGCS RGTATATSK TCAATAGGG AACATTTGAG GAAGTTTGT CTTTTTTTCA 420
TCGATCACA CAATCTCTT TAGAATTTGA CAAGTAAAT TTCTGTCTTT TTCAGAGACAG 480
TAAAGAAAG GGTAAAGAC TGAAGGAAT CTGCAATCT CAGTGTGAC GCAAGGATGA 540
TGTCTTTGAA ATTTTAGTGG AACTCTTCCA AGCACTGTT TTATGTTTAT ATGATATAAA 600
TAGTGTGCA AATATATAT GAAATATGTA TATAGTGTGG CAGCAATTT TACTCTGAT 660
GCAGTGTGAG ATTTGTATTT CATTGTACT ATTGTGTTC TTAAGATAT CCAATCTTTT 720
ACAGAGCTT TTGCGAATA CTTCAGGGGG CAAACCTGT ATGTCTCTTT TCGGAGKST 780
ARATTTTGA ATATATATCT RMTKSNATY AVGTGTGTA CNGEAMATA TTAATATTTA 840
TACATATAT YMTTATTSK KPAKATAT CCGAATATGT GGAATATTA CTTGATATTC 900
AAATGAAT CCGTGGAAA TTCCGCGAGG CTCACGAGG TAACTGTGAA TCTGAGCTTA 960
CCTCTGAGG TTAATATTA GAAGCCCTTA GTGTCCGAC AGTGGAGCAC AATATTCAG 1020
AACTTAAGA TATATTTCA GACAGCAC TCAAGCTCT TAAATGCTTA TCTCTGTGAC 1080

CTCTAGTAT GGGGACATC AATTGATA CTCTGGAGG AACACTGCT GACATGATA 1140
GAACTGACTT ACCCAATCTT GACACCTGCT CAGCTGAGCT TCAATTGTTG AGAATCAAT 1200
GGAAACAGG GGGGAAGAT ATGAGCTTC CTTCACAT CTATGAGCC CTCCACCTGC 1260
CTGACATCA GTTTTCTCT AATGTATG CATCTGAA GCTCTCTCT ATCTCTCTG 1320
TGTATAGGT TGTATATG CGGTATGAA ATGAGGAAA GGTCTTAAA GCAATTTGA 1380
GCAACATTT GACAGACCA AGTCAAGTA ACTTGACTTT GCTTAACATA AATTTTGATA 1440
TAAACACCA CTCTGATTTA ATGTGAGCA CATTATTTA ACTCTATACR AATATGTAG 1500
HCTTCTAC AKATATAT SHAKTGTG AATWSSSTA KESWGWCK TAWBRWKG 1560
GAWTWTYK RWTGATWY YGZWACAG AAAAACTT AGGTGTATG TAGACCATTT 1620
AATCAATA TATCTTTGCC TATGAGCTC CATGTATAC ATTAGCAAT GATAATCTAC 1680
CTTTTAAAT GGGCTCTGTT TGAATCTCA AGCTTTGAG ACTTACCTCT TCTTCAGAA 1740
GAGAACCTTG AAGTGCAT GTTTCTTTTT GGTGTATCT TTTTATGAG ACTCTGANT 1800
TGTCTTCAAT TAAATCAAT TTAGATATG CATTATTTAT CACTGTGAT CTCTACTTGT 1860
TGGTGTAT GAACTCTTTG AAGATATAT TTTGAGAGG TGTGGAGGA AGGAATCAT 1920
TTTATTAAT GTTCTATCA AGCCACAT TGACTTGA CTAATAGAG TTTTATGAT 1980
GTTAAATAT TATCTGAC AGTTACAGA AATTACCGA GAAAACTTG TGAATCAC 2040
AATCAGAT TCAATGTAG ATTTGTCTT TCTTCACTT AAGAAACAA ATGCAAGT 2100
TTGATGAAA AAGCTGCTG TTGTTCACA TCTCTTCTT GTTTACATC CTTTGTGAG 2160
CTTACTCTT CTATACCTT TTAGCAGTA TTGTGTGAC ACTTCTGTT CATGTGTAG 2220
ACGATATG AGCCATGTA TACTGACAC TGAATTTCT GTTTTCTT TCTGTCTTT 2280
TCTATGACT TTATATACT CTCTACTTG ATTTATAGC AAAACTGGA AACTTACA 2340
AATAGTGT GTGTATATC TGAATAATA TGGAAATAT TGTCTTTT TTGTGTGAG 2400
AAATCAAT TTGTATGTT TATTCATC TAATTAAT GTGAATTTG TTAAATTAA 2460
AATWGSAC AAATBHG GGGTCCAAA CHTWTCGS KAATTTCTT WAARATYK 2520
ATAACACT TCAATTC 2539

(2) INFORMATION FOR SEQ ID NO: 110:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1751 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

5 AGCATGAGC GAAATGCCCT GATGGCAGT ACCGTCTCG GCTGTGTCCA AAGATGCTT 60
GGTTTGGCG GAACTCTGCT GATGTCTTAC TACGTATTTG CATTCAATG GATCAACTTG 120
TTTAAAGGG TCAATGTGCG TCTTCTGGA AAGAGAGCC TGGCCCTCG CATGTGCTCG 180
GGGCTCTG GAGCTTCCA GAGCTGAGG TACTGGGCA ACMACTTGA TGAATTTGCG 240
GCTGCTCTG TCACTCTG GAACTCTG GATGTCTG ACTGTGACA ACTGCAAGT GTTCTGTGAT 300
15 GCAATGCGC GCTACTGAGG CCGTGTCTC AAGTCTAAT TTGTATTTG GTGCTGTGCT 360
TGTCTGTGA TCTGGTTCAA CTTGTCTG GCTGTGATTC TGGAAACTT CTTTCAAG 420
TGGAGCCCG GAGGCACTT GAGGCCCTT GCTGGAGCC CAGAGGCCAC CTACAGATG 480
ACTGTGAGG TCTGTGTGAG GATATTTCT GAGAGCCCG GAGAGGATGA GCTCAAGAG 540
AGGCTGAGC AGCAAGCCGA CTTGTGCTG TCAAGGTGAC GTTGGGCTG CCAATCCAGC 600
25 AGGGCGGCA GAGAGAGAG GCTGGCTTAA CAGAGTCCC CATTATGAA GAGGCGGCA 660
TCTGTGCGC AGCAGGAGG GAAGGAGCT TTCTCTCTG GAGCACTTAA GCTGGGAGCA 720
GAAACAAAT CTTTGTGCTG TGGCCACAA ACCATCTACA GAAAGCTGC TGTGTCTTCA 780
GGAGCGGCG GTGCCCTCG CTTTCTTTTA TACTGCTTC AGTGAATTT CCTGTGTGCA 840
CTCAAGAGG ACTTTTCA GAAATATGA AAGAGAGCG CCTTCCCTG TCTCTGTGAG 900
35 CTTGTGCTT GCTTTGCTG CCGAGCCG TTGGGAGCA CAGGCTGAC CAGGCGCTGC 960
AAGGTATAC GGTGATCTG TCTATCTAT TCAAGCTG GATGATACT AATACCTCG 1020
ATTTAGCCG AGCAACAG GGTAGCTTC AGTTTCTCT TCTTTCTATA GCTGTAGGC 1080
40 CTTTCTGCG AATGCTTCT ATCTCTTTA ATCTATTAAT GGTTCAGTTT TCTGTGATGT 1140
CCCAAGCTC CATTCACTG CAGCACTG CCAAGAGAT GCGCTGCTCA TCGAGCTGCG 1200
45 CTTTGTACT CCAAGCTG TACCCCTCTT GTTGGAGCC CTTGTGCTCA AAGCTTTGAG 1260
CAACAGCTT TCCAAATGA AGTTGTACT GTCAAGCTT TACTATCAGC AAGCAAAA 1320
TCTATAGCT GGTGAGGCTC CTGCTCATT AAGATGCAAT AATATGTA GTACATAAA 1380
50 ACAGCATAG AAGAACTTA ATGCTTATT CTCAAATAT ATGTCTTACT AAAAAAGCA 1440
AATATATTA GATATGTAAG AATTCACCA GCACTTTGG AGCCCGAGCC GGTGTGATCA 1500
55 TGAAGTACG AGATCAGAC CATCTGCTT AAGAGGTGA AACCCCTCT CTACTAAAA 1560
TACAAAAAT TGGCCCGGCG CAGTGGCGG CCGCTGTGCT CCAAGCTACT GGGAGGCTG 1620
AGGAGGAGA ATGGCTTGA CCGGAGAGC GAGCTTGA GTAGCCGAG ATTGCGCCAC 1680

TGCATCTGCG AGTCGACCT GCGCGACGAC GCGAGACTCC GTCTGAAAAA AAAAAAAAAA 1740

AAAAAAAAA A 1751

5

(2) INFORMATION FOR SEQ ID NO: 111:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1117 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AATGTTGTGG TGGTGGCAAT TGGGTTAAAT CTAATATAG AGTCTCTTGG AGACCAATGT 60

CGATTAACCTA ATCCGACCA AATTGTTCTT TTATATGTGG TAGTGACGAC CAGAGAAATTT 120

CAAGAGCCT TCGTAAATATC TGTAAAGCGG CAGCTACAGC AGGCATCAAT GCGTGGGTTT 180

ATGGGGGAAT ACCAGCTTTT ATTCACTCTA ACGACAACTA CATTGACGAC ACCCGGCGAG 240

AAATTTATCA TAAAGCGTTT GATCTGTGCG AATCTGCACA TCGTCTGCGC ACAGAGGCT 300

TCAATCTGTA TCGCTGTGCG TGGGTTGGA GAATGCGAGT GTTTGTGACT ATATTCACAA 360

CAGTGAACAC TAGTCTGAAT GTATACGAAA ATAAAGATGC CTTAAGCCAT TTTGTAAATG 420

CAGGAGCTGT CAGGGAGT CTTTTTNGA TAAAGTTAGG CTTGCTGTGC CTGTGTGCTG 480

GTGCGAAT TAGAGCTTGG CTGGGCGTCT CTGTAGAGCG CTTGCTGTATG GCATTTTCAGA 540

AGTACTCTGG TGAGACTGTT CAGGAAAGAA AACAGAGGA TCGAAGGCC CTCGATGAGC 600

TAAAGCTGGA AGAGTGAAA GCGAGACTAC AAGTTACTGA GCACCTGCTT GAGAAAAATG 660

AAGTATGTTT ACAGGAGAT GAATCTGAGA ATGATGCTTA GAAATTTGAA GCATCTGCTAA 720

ACCTTCTTAG AAACCTTTCA GTAAATGATA AACAGACAA GAACTGAAG TCGTCTGAAAC 780

TTGAACTGCA CTGGAGGCT GAAGGAGCT GCGATGTCCG ATGATATCCA ACAGCAGCGC 840

CACTCTTTGG TCAGCTGCTT GACAAATTTA AGTGTGTTA CTTGTGTGCG CAGTGGCTTG 900

CTCTGTGCTT TTTCTTTTCT TTTTACTAA GAATGGGCT GTTGTACTCT CACTTTACTT 960

ATCCTTAAT TTAATACAT ACTATGTTT GTATTAATCT ATCAATATAT GCATACATGA 1020

ATATATCCAC CCACTGAT TTTAGGAGT AATTAACAA TTTGCGAAA GATTAAGTTT 1080

GAATTTTACA GTTAAAAA AAAAAAAAAA AAAAAA 1117

55

(2) INFORMATION FOR SEQ ID NO: 112:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1654 base pairs

(B) TYPE: nucleic acid

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1313 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GGCAGGTT TCTTATAT TTATGTAAT TTAAATGCG TATGCAATA TTTATCTTGG 60

TTTGAGACTA CCAACATAAC TACGTTTGA AGTGTCTTCA CAGGAAATAT ATTGCTTTTA 120

ATGTCAATA ATTTCACCA ATGTGCTTA CTTTATAAA GTATTAATTT TGTAAATAT 180

TCAATTAAT AGTTGTAAAC CTTTCTTAT TTATGTAATA CTTAATGCAAT GTTTACTTTT 240

TTTGAAGA TCGACCAAT CTCCTTGAC ATGAAATGCG GCGTCTCAAT ATGCTGCAAT 300

TACTAGAAA AGAAGAAAGA CGAATCTTC TTATGTAATCT TCGCCATGT TGTAAAGGCT 360

TGTGCTTCA AGCGATCTT CGAATTTGAA AGCAGCTTGT TAAAAATGAG AAGGGCACCA 420

TAAAGAGC TTACAGAGT CCGTCAATG TAAACATAA ATTAATTTGA TTAGTCTTTC 480

GATTATTTA GCGAGAGCT ACTTGCATG CTCGAGACA TGAAGAGCT GAAGATATA 540

AACTTTACA GTCAATATC CAAAGAGAC TGTGTGTGTC TTAAAGACCA AGTTACATA 600

TGTATTTTT GTTACTGTCT TCGTTCAGA GTCAATATC TTTTGCAGAG TTCTTTGCTT 660

TGACAGCAT TAGTGACAAA GCGAGAAAG ATTATACAG CATGCTAAAA GAGTGAGAA 720

TTTGTATCT TAGAGACT AGTTTGTGCG AACTTAAGAT TTACGTTTAA TTTTTCATA 780

GTATTTGACA CTCATGAAA ATAAATGAA AACATCTAGA TTTATGTAAT TATCTGTGCG 840

CTTTGTATA AACTGAAGT TTTGAAAT GTTGTGCTCT GCTCTTCCAG CCTATGATA 900

TTTTTTGAA ATGGAACCAT GAATTAAT CTGATCATC CATACAGAC CAAACATTTT 960

AATCAAAAC AATGTGTTCA TCAAGTAAT TCGTCAATTT GTGCAATCT ATGTTGTACA 1020

GACCACTGA AAGGAAATGC TGGTCTAGCT GCGTGTGTA GTTTATAGCG GAATTTGAGC 1080

AGAGGAGCG CAATATGCT TTTTCTTTT GAAGTTTTT TAAATATAT TTAATGCTTC 1140

TTTTTTTAA TTAAATATG TGCATTTTA CAATTAAT TGGATGCTCT TTTGACCTTA 1200

AATGCTTTT TTTGATCAG AGATGTGTA CTATTTTAT TTTTATAAA TGTATCTTCC 1260

CTTTTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA 1313

55

(2) INFORMATION FOR SEQ ID NO: 113:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1654 base pairs

(B) TYPE: nucleic acid

60

336

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
5  ACGGAGACG AATACCTTCT TTCTTCTCTT CAGTACAG AGGCTTCTT CTACACTTG 60
   CGTCTACCTT TTAATTAA ACCTATCTT TTCTAGTAGT ATTATCAT GCGATGCGA 120
10  TGATGACAC AACAGCTTCT CATTACAGC TGAAGGAG CAGTCTCTTA CTAAAGTAG 180
   TTCTCTACT TTCTCTCTA TTATAGGA ATTATACGA TTCTAATAT ACCTATATT 240
15  TTCTTTTAT TTATATAC CAGTACAG AATCTTTTCT CAGTCTTCTA GAAATCTCT 300
   CATAGATCA CAATATCTA CATTTCATTA GATGATATTT TTCTAGAT TTCTAGATG 360
   CAGTGGCAGT CGATGGCTTG GTTGAAGCTA GAAATTTTCT TGGCTCTCTT GACCTGTAA 420
20  GCTCTCTCT CCGAACCTTG TGATGCGCTG AGGAGATGA CAGATGCTCA GATGAGAG 480
   AGAATATAT GAATCTCTCT GCTCTCTCTG GTTCTCTCTT TGGCCGAG TTTTTCAAA 540
   GCGACGGAA TTAGACTACT TCAGATCTCT ACCTTTCTCA GCTTTTCTT TATTTTCTG 600
25  TCTTTAGAT TTCTCTCTCT AAAGGGCA AGAATATTA CTCTCTCTT CTACAGACC 660
   ATGGCAGCTT ACCGAGTAG TCTTTATGCT GGTGTTTAGG CATATAGAT GCTGATCAT 720
30  AGTCTCAGC CAGATTTCT TTCAATATC GTCCAGCTTG AGTGTCTCTT TCTCTCTCT 780
   CCGATTTCTT TGACCTCTCT GCTCTAGCTT TGGCGGCGG AGAGTCTAT TTGCTTTCT 840
35  TCTCTCTCTG TCTTAGAA ACCCATCTTT AATATATTC TTCACTACTG TTGGGCTTCT 900
   TTTGTGATTT TTTTCTCTT CCGAAGACT CCTGCTTCTT ATTGGATTTT GTATTTTAT 960
   ACAAATATTT GAATTTTATA ACCTCTTACA CATATTTAA TTAGTGTGAA AGGAAAGAA 1020
40  GAATCAGGA AAATATATTT AATATCACC TCAGTTGACA AGGTGCTCAG ATTATTCAT 1080
   TGGGATCTCT CTTTCTGTTA GCTTTTATAG ACAACCTTAG ACTTAACCTG TGTCAAGAC 1140
45  TTCTGAATCT TTAGCAGTG CTAGTAATTT CCTCTGATG ATTCTGTTAT TACTTCTCTA 1200
   TTCTTTATTC CTCTTCTCTC TGAGATTTAA TGAAGTTGAA ATTTAGCTTG GATTAATACA 1260
   AAAGATAGT GTATAGTAT AGTATCTTAA GTTCAGATGA AGTGTGTGTA TATACATCA 1320
50  TTCAAAATTA TGAAGTTAG TAAATCTCA GGGTTACTTA AATTTACTTA ATATCTCTTT 1380
   GAATCTACTT TGTCTCTCTG CTAGAAAAA TTAATACAG GACTTTGTAG TTTTGGAGC 1440
55  CAATATGATA ATATCTATG TTCTAATAT TGGGCTATAC AATATATTT AAGAATATG 1500
   GATTTTATTT CCGGATAT GGTGTCTCTT TTATGATAT ACCGAGATG ATGTATGAG 1560
   TAAATACAT TTGTAAATA TGTAAATATG TCAATATTA AATATCTTT GACTTATTC 1620
60  CAAAAAATA AAAAATATA NTTGAGGG GGGC 1654
```

337

(2) INFORMATION FOR SEQ ID NO: 114:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
15  GCGAATCTTT CCCCAGAGC TTCCAAACTT GCAAGCCGAA ACCTTGATC GTTAAAGTT 60
   GGGTTGCGAC GGGGCTCTGG CCCCAGAGAG GCGAATTTGG GTTCCGCGAA CGTTGGCCCT 120
   CAAAGCTCTG GCAAGCAGCC ATGCTCTGCA CCGAGGAGAG GGGCTCTGG CTACAGAGAC 180
   CTGAGCTCTA TCTTCTCTGG CCGACTCTGG GGGGTTNAG GGAATTTTCA GACTGTGAG 240
   GAGTCTCTGC TGGACTGCTT GTTGGACTTC TTACCCGAGG GGTGAGCA AGAGAGATC 300
25  AACACACTCA GCTCAGGA AGCTTATCT CAGAAATATG TTAAATGCT CAAATGACTCT 360
   GACCGATGGA GTCTTATATC CTTCTCAAC AACATGTGCA AAATGTGGA ACTGAAATTT 420
   GTGATATCCC TGGGAGGCA GTTTGAATTC AGTGTAGATT CTTTTCATAT CAAATTAGAC 480
   TCTCTCTCTG TCTTTATGA ATGTTGAGG AACCCATGA CTGAGATTT TCACCCGCA 540
   ATATCTGGGG AGAGCTCTTA TGGGATTTTC CAGGAGCTT TTGATGACT TTGTAAAGAG 600
35  ATCAATGCGA CAGGAGCCC AGAGGAAATC CGAGGGGAG GCTTCTTAA GTACTGCAAC 660
   CTCTTGTGTA GGGCTTTTAG GCTCCCTCTT GATCAATCA AGACCTTCA AGGTATATG 720
   TGTTCAGGTT TTTTCATGCA CTTCTCAGAC ATTGAGAGC AGCAGAGAA ACTGGAGTCC 780
   TATTTGCGAG ACACCTTTCT GGGATTTGGA AGAGCGCAG TATGATATC TCATGACCTT 840
   TCATGAGTG GTAAATGGA GCACAGTCTG CCGATGTGGA CATGAAAGA GACGACTTTT 900
45  AAACCTTATC ACATCTCTG CTATCCGGCT GTTAGCTGAC CAATATGTA TTCTTAATCT 960
   GCTTAATCTC ACTGCTATTT ACCAGCAGC CCGCTATGTA GCAATGCGA ACTTTAGCAA 1020
   TTACTAATTT GCAAGCTTTC AGCCAGTATT CAGCTGCGAG CAGAGACTT ACTCCACTTG 1080
   GCTACCTCTG AATTAAGAT CATTTAAAAA TGTCTCTCTG GGAAGCCTT TCAGACAGA 1140
   CAGGAGAGAA AAAAAAAAAA AAAAAAAAAA A 1171
```

(2) INFORMATION FOR SEQ ID NO: 115:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 842 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

5 GGTCTGCGCC GGAATGCAAT GAGCTGCGCA TGTGTGCTT AGTGAATGCG GTTTCGTGCG 60
10 CTCTCCGCGG TTTCCTCGCGG TGGGTATTG CTTGCAACCA TGGCGGCCGA GGGCAAAATG 120
GGCAGCAGAG GGAAGAGCA GATATTGGA GAGACAGAG AGACTCTGA GTTCTACTCG 180
GGATCTATAC TGGGGGCCAA TGCATTTAC TGCCTTTGGA GGTTCGTCTT GTTTTACTCA 240
TCTGCTCAAT TTTCGGCGCG GTTGGCGCG TGCCTTTATG TGCAGTCTGA TGGGGCCAGC 300
TACCCTCTTA TGAATCTGAT GGAAGAGCA GCGTTCTCTG AGGATGCGGC CTTGATGAT 360
20 GGTGCAATCG ACCTCAACAT GAGCAGAGC ATGCAAGAGC AGCTTAAGA TGTGATCTTA 420
CTGACAGCA TGTGCAAGT GCTGAGTGC TTCTCTCTCT ATGTCTGCTC CTTCCTGCTT 480
CTGGCTGCGG GCGGGCGCTT TTACTCTGCG TGGGTGATG TGTGCGGCC CTTGCTTCTCT 540
GCGACAGTGT GACCTCCAGC ACCAGAGAC AATGAGAAC GGCAGCGCGG ACGAGAGCGG 600
GCGACAGTCA AGCGTTTATA GCGATTGCA TTGTGGCCAC AGGCCTACTG CCGTGGGTGG 660
30 CTCTCTGAGG GTGCAAGCC CTTGATGCTT GAGCAATGA GGGTCTATGTC CAGGGGCCAA 720
AAGCACTGCT AGGTATTGCG TATACTTATA CTCTATPAGG TGTGTGATA AATGCTTTAG 780
35 AATGTGAAA AAAAAAAAAA AAAAACTCG AGGGGGGGCC GGTACCCCAT TTCTCTTAA 840
AT 842

(2) INFORMATION FOR SEQ ID NO: 116:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1640 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

50 GGCAGAGCG GCGGCGAGCG GTGCGCGCGG CCGCCCGCGG CCGGAGCGCT TCCCTTTCCG 60
GTGCGGAGCG GCGGGGTGCG GCGCCAGCGG ACCCGCGCGC AGGAGAGCGG GGAAGAGGAT 120
GGATTCGCGG GCGCTCGCGG CCGCATGCGA GAAAGAGGAA GTGATCTGGA AATCTGCGCT 180
AATCTGCGG AAGAGCGATG TCTACTACTT CAGTCCAGT GGTAGAGAGT TCGAGAGCAA 240
60 GCTCATGTT GCAAGTACC TGGGAATATG TGTGATCTC AGCAGTTTTC ACTTCAGAC 300

360 TGGAAAGATG ATGCTTAGTA AATTACAGAA GAAAGAGAG AGACTCGAA AGATCTCTCT
420 CAATCAAAAT AAGGTAAAC CAGCTTTGAA ATACAGATTT GCAATTTAGA CAATCAAGCT
480 CAATTTTCAA ACAAAGCTTA ACCCAAGTTC ACUAATCATC CTAGTAAATA ACTGAAATCA
540 GACCCACAC GAATGATACA ACAAAGCTT CAGCTTTTCT GGAAGAGAG GCTACAGAGA
600 CTTTAGTCCA TCAGATGATA CAGAACAAT TATATAAAAC ATGGAAGTAC CCAAGGTCTT
660 TCAAGAGATT GGTCCAGTAG CATTGATGAG AGCCTTTTAT CTGTGTGTC CAGTCTTTTG
720 CAAGAGCT CTGCGCAAT CAGAGGGA GTCTCGCGTG CTGTGAAA GAACTGCTG
780 TTGTGCTTAA CAGATCTCAA CCGCTCTGCA AGCTTTTAT TTGTAAGAT GAAGACTGAG
840 GAAAGAGAG AGCAATGACA GCAATGAGC AAGAAATGAG AAGAGCACT GATGCGAC
900 ATCTGTGCG GAGCTGCTGA TACAGAGAG ATGATATG AATGAGAG TGGAGATGAA
960 GCTTAAGAT ATGATCAAGT AACTTTGAG CAGCTTTGCC CAGAGAAA TTCTTAGA
1020 AATTGACAA AATGTTTCCA CTGCTTTTG CTTGTAGAA AAAAAATGA CCGAGACAA
1080 TAGAGCTTT TAATAGCACT AACCAATGCC TTATTAGATG TATTTTGTAT GTATATATCT
1140 ATATTCAAA AATCAATGTT TATTTGAGT CTTAGAGATT AAATTAGTCT TTGTTTAA
1200 TCAAGAGA CCGTAAGATG AGCTGAGCT TTGATGCGA GGTGCAATCT ACTGGAATG
1260 TAGCACTTAC GTAAACATTT TTTTCCCCC AGAGTTTAAA TAGAGAGAG TCAGGAATTC
1320 TAAATAATT TCCAGTTTAA AGATTAATGT GACTTCACTG TATATAACA TATTTTATA
1380 CTTTATTGAA AGGGGACAC TATACATCTT TCCATGCTCA CTTTAAGAC AATTAATGA
1440 TTATATTTCA CAGAAAAAA AAAAAAAW WISTYGARER GSRGCKELW AYHWRWOC
1500 CCGGRTWRGS MTCSTWKA YTTAGATTA ACTCTGATCC CCGGCGCTTA GTTTTCAGAT
1560 GGGAGTGGG AGGAAGATG CCGTATATT TCCAGTATGA ACTATTGCTT CTGCGAGCTT
1620 GTGAGAAAT GTGCTTTCAC CAGATTCTT AAGATTTCT GCTTTAATA TCAGCTAGCC
1640 TGTGTAATT TTTTTCCTT

(2) INFORMATION FOR SEQ ID NO: 117:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 952 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

60 TGAATTAGN AAACACTTG GAAACTCAT AACCTCATCA GAACTGCTT TTAGCCACAC

5 TCTGACCTT CTAGATGAT AGCAAAAAA TGAATATAGT TCTTGGAAAT TAGGCATTT 120
ATTTTAAATT GCTATTTTTT TCAATGTCTT AGATATCTTT AAATTTGTTA TTGTGGATC 180
ATTTTCTCTC CAGATACCTT TATCAAAAT ATTGGCTTCA TGGAGCTGA AGTATGTGAG 240
CTTTTGTGTC AACTTTAGTG GACTTCTGTG AGATGTAGT TGTACTTTGT ATCTCTAAAT 300
CTAAGATAG TTTTATAAA CTCCAAAGA AAATCTGTC TCTTTTCTCA TCTAAAAAT 360
CATCTTTGG GTAAAGATTT AGTCTTCAA AGTTTCTCAC AGTTCTAGAG CTCAGAGGGA 420
GCTAGCTCTG CAGCTGACAT CTCTGCATCC AGAGCTGACA GATTTCCACA GAGTGTAT 480
TAAATCTCC AGTATGACAT GCTGTGTTAG GAGGGGGTGA GGGCTGGCTT ATTATGATAC 540
AGGCTGCTGT ATTTTACAT GTTWTGGG GAAGGGGAGC CTGGAGAAA CAAGTCACT 600
ATTCCTTTT TTGAACAGG AAAAAAAT ATTTTGTCTT CAGTAAAT GTTAGAGAT 660
TCCATGTCC CTAGCCACA GGGACAGTT CCACTGAGAA GTGAGCATG GGAATCTAAA 720
ATTTGAGAA CATTGGGGA AGGAAAT GTCTTCTCT TATTTGGAG ATGTCTCAT 780
GGGSGGGG GGGCTGTGTT TTTTGGAT GTTATATGTT GTATGTAGC ATATATGAC 840
GGATCTCC TGAATTTATA AGTTCCAAA ATATGATTA ATCTTGTGTT TTTGTAAAT 900
TATCTCAATA AAGGCACT GBACTTCAA AAAAAAAA AAAAAAAA NN 952

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 118:

5 GAGAGTGA AGACTTCTT GGCAAAAT GGCCAGTTTC AGAAATAGT TGTGTGTTTA 540
ATTGAGCTTG TTGATCACT TCCAAAGAA GCGAAGAAATG AAGAGATGAA GGCCATGGT 600
GCTGGAACT TGTCTAAATC TATAGCAAG CAGAGAGAG CTCACAGCA GCACTTCAA 660
GGCTTAATAG CAGAAAGAA AATGAGCTA GAAGGTATC GGTGTGAATA TGAAGTTTG 720
TGTAAAGTAG AAGCAGACA AAATGATTT ATTGACCAAT TTATTTTCA GAATGAACT 780
GAAATTTTG CTTTATATGT AGGAAGGCA AACAAAAAA AGCTCTCTCA AACAAAAA 840
ACCTCTTAG CATTCAGG GCTTGACCA TGAATATGT CACAGAGGT GGGTGTAG 900
GAATGAGCC CCTGAAAGAC AGCACTACA GTCTGGGGGA GCGATTTTA ACATCATCC 960
ACAGCTGCTG CTGTGGGCC TCGAGGTAC GTTCTCACT CTTATGCTTA GTTGAATTA 1020
AGCAATTTT AAATTTTAT CTTTTTTT GTAAATTCAC AAGCTTTTG AAGAGAGCC 1080
AATAAATTT TGTTTTCAA TGGTTGAG TACTTTTTT CTTTGTCTC TTCAATATG 1140
TTTAACTCT CATGAGAA CCTTGATTC TGTATGCTT ATTCACAAA ACAAGCAGG 1200
CACTGTGAG CAGTACTT TAAATGAT CACAGCTG AGTCAGAG TACAC 1256

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 119:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

5 GGGCTAGCA GCGGCGCTGG TCTCTCTGG AGCGGCGGC CCGAGTGGG GCGCGGCAT 60
GTACTTCCA CATTGATAT TCGAAGAA GTGATCTGA CTCGTGACAT TCTTATGCA 120
TACTTAAT CAAATATAG AGTCTACTA CTTGACAC TGGCTCAGC AACATGAC 180
GTGGAGTTG CCGACATGA AGTGAATCA ATATCCCTTG TATGAGAG CCGGGTATG 240
TGGTCAAT ATCAATGGG AGTTGGCTTG CTTATATG TCTTACTAG CATTCCTTC 300
TTCAATGTC CTGTGCTTG GACTTTACA ATATATATC ATATCTGGG GATATGATA 360
TTTTTGATG CAGTGAAGG AACACTTTC GAATCTCTG ACAGAGTAA AGCAGGCTC 420
CTACTCAT GCGAACACT GACTATGGA GTACGTTTA CATCTTCAG GAGTTTTC 480
AATATTTCT CAAATATCT ATATTTCTG GCAAGTTTCT ATAGAGATA TGAATCACT 540
CACTCATCC TAAACAGAG TCTCTCTG AGTGTACTAA TTCCAAAT GCGACACTA 600
CATGTGTC GATCTTGG AATTAATAG TATTAATG TTTTGAAT TTTTGAAT 660

TTTACAGCTA CTGAATTTCT TTTAGAGAG GAGTGGTTAG TAAACTGCAC TGTTCCTSTG 720
ATAATCGAA ATGAGAGATA TTTACATTCG AGGCGCATG GCTGCTGCTT CAGTCTCT 780
TTTGAGGTGC AGATTTCAT TAAATGATGC CTCGTGTTAA TACACCTGCT AGATTCTCA 840
AGAGGGGCTT TATAGCAGG CTGGCAGCG CCAGCTTATA AGTTAAGCG CATTACAGTG 900
AGGCTGTAGT AGATTAATTC AGGAATATA GAGATTGTGA AGAACTAGG ACCAGCTTAA 960
CTTATATCA ATGGCAGTG TGTTAGGAA AGACATTTTC CAGTCAATCA GCTGTGTTTA 1020
TTTAAAGCAG ACTTACATCT AAACCGAAT CCTCTCTATA CAGTTTATTT AAGATTAAT 1080
TTTATTACCG TAAAGAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1140
GAG 1143

(2) INFORMATION FOR SEQ ID NO: 120:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1782 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

CAGGCCCCCG CCCCCACC AGTCTGCT TCTGCCCCG CTTGGGCTCG GCCCCAAGG 60
CAGGAGAA GAGCTGTCA GGAACCTCC GCGGAGTGC AATTACGTG CAGCTGCCG 120
CAGCCAGG TTTCCAGTG GTTTGCGGG GCTTGGCGTG TTCCAGGAG TCGCTGTGCG 180
CCTCAAGCT GCTTTACAC TTGGTTAGT TCTGCTTAAT TGGAAATGCT GCTTGGGCA 240
TTGGCTTGG GCTGATTTCC AGTCTCCAG TGTTCGGCT GTCTATTGCA GTGGGCACT 300
TCTTTGCTT GATTGCTTTA GTGGTCTGA TTGGAGCTGT AAACATCAT CAGGTGTTC 360
TATTTTTTTA TATGATTAAT CTGTACTTG TATTATGCT TCAATTTTCT GTATCTTGG 420
CTTGTTAGC CTGAGCAG GAGCAGAG GTACGTTCT GAGGTGTGT TGGAAATTA 480
CGGAGTGC TCGAATGAC ATCCAGGAA ATCTAATCT GTGTGGTTC GAGATGTTA 540
ACCAATAGA CAGCTGTCT GCTAGCTGT TTAAGAGTA CCAGCTGTC TCGCATGTG 600
CTCCATCAT AGGATTAAT GCTGAGAG GTTTGAGAT TGTGTGCG ATTTGCTGT 660
TCTTCAGTT TACAGATCT CTGGTGTTT GCGTACCTA CAGTACAG ACCAGAAG 720
ACCCGCGCG TATCTCTAT GATTCCTTT GATGAGAAA CAGGAGAT TCTCTTCT 780
ATTATGATC TGTTCACCTT CTGTAATTT CTGTTAGCT CCAATTGCA GTTTAGGAA 840

GGAACACTA TCTGGAAG TACCTATG ATAGTCGAT TATATATTT TACTCTATGT 900
TTCTCTAAT GTTTTTTCT TTTGCTTCT GAAAAATAT TAAAACTTGT GGTCTCTGAA 960
GCTGGTGGC ACTGGGAT TTACTGTAT CATTTGCGG CACTGTGCAC TGTGGCTTT 1020
CTTACATTT TTAAGTCAG AAAAAGTTTG TATGTACCA CTGTGTGCT TATATGTTGA 1080
ATCTGAGCT ACATCTCAT GTATATATA TATGTAGAC TGTGCTGTGT AGATATGTC 1140
TACTGAAAA AGATGTRAA TTTATTAATA TCGAAGATA TCGATCTG TATATTTAG 1200
GGAATCCAA ATTCGCAAT TTTTGTGTC TTTTATGAA AGATGTGTT TGTAAAGAG 1260
TGTATGATA AAATGATTA TTAAGTCTA GTCTTTTATG ATAGACAAA TGTATCTAG 1320
AAATGATTA GCTTAGGAA ATGTGCTTT ATTTTGTAC TTTTACAGT AGTGCAGAG 1380
GAGAAATGT TTAGGAAT GTCTAATGT ATATATCAT TTAAGTTCAG CTTGCTCAG 1440
AATGAGCA GTTTGAGTA ATCGAGAGT ATATCATAT CATCTGTATA TGTTTTATA 1500
ATAATTTGA GTCTAAGAA CTGCATTTT AAACAGTTA GTATTAATGC GTTGGCCAC 1560
GTAGCAAAA GATATTGAT TATCTTAAA ATGTTAATAT ACCGTTTCA TGAAGTTCT 1620
CAGTATTTA ACAGCACTT GTTAAGCTA AGCAATTTG AATATGATCT CCAATATTT 1680
GAATGTAAA TGTATTTG TGGCTGTGA TATCTGTGA AAAAAATTA GAGAGAAAC 1740
CTTCTTTGT GTATGCATGT TTGAATTA AGAAGTAAT GG 1782

(2) INFORMATION FOR SEQ ID NO: 121:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 610 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GTTCGCTGCA GATTGTGCT GGTTCGTGAG CGTCTGTCC TCGCCAGCA TGTCTCAAG 60
TATATTAAG AACATATGCA TCCCATGAA GCTCTACTAC ACCAAGTTT ACCAGAGAT 120
TTGATAGGA ATGGGGCTGA TGGGCTTCAT GGTTTATAAA ATCGGGGCTG CTGATTAAG 180
AAGTAAGCT TTAGAAGCTT CAGCGCTGC TCTGTGAT CAGACCGCA TTTACTTGA 240
GTACATGCA AGAAGAGCT CAGTCTGCT GTAAATTTCA GCAAGCTGT TTAGATGGG 300
ACCGTGAAC GTCACTGTAC ACTGTATTA GTACCGTTTA CTTCATGCA TGAATTAAT 360
GATCTGAG ATGCAGTCT ACTGATCAT GCTTTCAGT TGTTCCTCT CAGCGCTCG 420
CGGTGTGAG CATACTCTGA GTATATATT TGTATGAG CCGATGCAAT CAGATCTCA 480

CTGAGCCACC CATTCTCTGT AATTATTTAC CTGATTTGTA CAGGACTTGG TCGATCGAT 540
CGAGGCACTC ACTTGTATTC TACTCTCTAA TAAAGTTTAA TTAACTTGA AAAAAAAAA 600
AAAAAAAA 610

5

10

15

20

25

30

35

40

45

50

55

60

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 526 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2081 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

TGTACCGGTC CGGAAATTC CGGATCGACC CAGCTGCTCS GGGGACATG CGGCTACGG 60
AGCGCGCGGT CTTTGGGCTC CCGACAGAGG GCGCGGGGGG CGCGGGGGCG CGTTCGGGCA 120
CAGTCCCGGT GCTCTCTGT TTTTCAGTCT TCGCGCGACC CTGTTCGGTG CCGACAGCGGG 180
CGGCTTACA GGTGCTGATC CAGAGTTTCC TCGGCTCTTA TCGGCTCTTA CCGGACATCC 240
ACCGAAATTT CTGTTCGGAG CTGTTCGGCG AGGATTCGGG CCGATTCGATC GACTTTGCCA 300

AGCGCTTCCG GGTACCGAG CGCTTCAGGG TCGCGCTCTGT CGCGTTCGAG ATCCAGCTCA 360
CTACCTCTGG AATCTTTACA CTTTCAGCA CTGTCTTTT CTGTCTGTAT ATCCAGGAAA 420
GTTTCAGACC ACCATCAGG TATTTTGGGG ATATTTATAG COTTCGACAG AGATTCTTCC 480
AGGGGGCGCG GATTTTACGA ATCTCTTTTA TTGTACACGA ACATTAACCT AAAGTCTCTG 540
CGAGCAGCGT TCGAGAAATT GATTTACAG GTGTAAACT GTTACTTCCA AGACCAAGT 600
TTTCATCTGT ATTACACAAA GTACAGCGCG GATTACAGA GATTTCCGGA GTACAGAGTG 660
TTGTATTTT TCGATGAA ACTCATGTGT GCATCCACA ACTGCGCTTG GAGCTTATTTG 720
CGGAGGAGT CGAGTTTAC ATTTCTTCTG ATCCACTC ATCAGAGAG ATCATTCACA 780
CGATCTTTCG CTTTCAGCGT CTGCTCTGAT CGGGATCAT ATGTACCCAG ACTGAGGCTG 840
TTCTGCTTCA GCTGTGAGT GATTAAGACC ATCCAAAT CAGGAAAT CAGAACTTAA 900
TTAGCGCAG TCTTCCAGAG TCGGCTCTGC TTTCCAAAGT ATAGAGAT TCGAGAGACTG 960
GTATCTACT CACTGTGAA GAGAGTCA GTAGAGACT GTAGAGCCAG ACAGCTCTTT 1020
CTTATCTCTA CTAGAAATTA AATGTTAAGT GAAAGCGCG TCTTTTITG CGGCTCTTAC 1080
TGAACTTTAA CGAGCTGAGC GATTTGAGTA CGAGATTTA GTTACAAAG TCAAGGCTTT 1140
CGGCTCTGCT TTACTCTCT TTTTCTTTTA TGTCTTTTA TTTATTTAA AATATTTACA 1200
TGAGAGTCC TTTTCTTCT CTACTGTGTA CTCTGATCT ATCTTTCOA AGTGCAGACT 1260
CTTTCAGAT TTTCTTAAAT TGTCTACTTT AAGAAATG AGTACCAAG AATGATTTGG 1320
CTTTTATTT ACTGTAGAT GTTATATGT TATGTGAT GTAGTCTTT TACTTTTACG 1380
ATTGANTGAA ATAGATTTAT TCGATATGAA TTTACCCACA CGAGCTCTGAA TCTATTTTACC 1440
CACTCGCTC AGATCTTGT CCACTTATG AGTTCATG AGTTCATCT ACCAAATGAT 1500
GTGAAATAT TACTATCTTT TCTTGACTAT ACTGATTTCT TATTTTGGTC ACTATTTACTA 1560
AATCTCTGTT AATATCTCT CTTTACTG AAGAGGATG GATTAAGAG GTTTGCAATG 1620
CGATATATT GGTTCAGGCG TTTTTCACA TCTTTGAGT ATGCTTGTCT GATATCTTTT 1680
ACCAAGTCT TATATATTA TCTGATGTC CAGAGATTT AGGAAAGAA TAAAGCTTGG 1740
GTGAAATAT ATTTTAAAT GTTATGTC TGTCTATAT TTTCTTACC TACTCTTCAA 1800
ATATTTAT GCAAAAGTC TCGATATGA TTTGTATGA TTAATTTGT GGTCTATGTT 1860
TCTCTGAT AATTTATTT TCAATTAATA CTTTATAG GATTTTGA TTTTTCAT 1920
ATATTTGAAA TGTGAGACT CTGTCTTTA TATTAAGTA ATTAAAGAA ATGTATTTG 1980
ATGAAATTA TTTTGGCTC CAGAGATG CTCTATGAT ATTCTTCCAG GATTTCTAT 2040
ATTTATTTA GATTAATTA TCTTGAAT TATATCTTT C 2081

346

347

5 (2) INFORMATION FOR SEQ ID NO: 124:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

15 CCCCAGCGGGA GCTGGACCTCG CGGTGGGCTA GGGGCGGGCG CGGAGCGGCG GCGGCGGAGC 60
TGTGGATCCT TCATGATGAG AGATTGGGG AGACTTCTCT CTCCTGTGTG TAGTGTGATAG 120
TTTGTGTGTC AGAGATGCG TCACAGTGTG AAACCTTTTC TCCAGACCT TCCGAGAGGA 180
ATCAAGAGCT CATCTGGGG TATTGTACC ATCTCAAGC TAGATGCTCG AATCCAGGA 240
AAGAGAGAGG AGCAGGCTCG AAGAGGGCA AGTAGTGTCT TGGCAGAGAG AAGAGCGGAG 300
AGTATAGAGC GGAGCAGGA GATGAGCCA GGTATGTGA GTAGATTTT CAGATGTGT 360
GCTTGGATG GTGAGTGTAT CTGCTTCACT CTCCTCTGTG TTTATCGAGT ATTATTCCT 420
GTGCTTCACT GGTATAGCG CCGATTATC GGTAGCCAT CACTATATG AGATGTGTG 480
TGTGTGCTCG AATCTTCTCT CAGTCAATT TTCAGTGTCT TTTGGTGTCT CCGCTGTGTT 540
GTGCTTAGCA AATGTGTGA TGGATTTGCG TTTGAGGATA TAGCTGACT GGCATTTGAG 600
GTATCAGGGA GGAGGCTCA CCGATTCTCT AGTGTAGCA AATATATGCG TCACATGCTC 660
TTCAAGCTTT TGTGTGAGCG TCTTTTCTTC ATTCAGGGA TGTTTGTGAG TCTTCTTCCC 720
ATCCATCTTG TCGGTGAGCT GGTATAGTCT CTCGATATGT CCGTCTCTTA CTCAGTGTAC 780
TGTATTGAT ATCTTGTGTT CAATAGAGA ATTGAAATGC ACAGGGGTT GTCTAACATA 840
GAAGGGAATT GGCCTTACTA CTTTGGGTTT GGTATGCGCT TGGCTTTTCT CACAGCAATG 900
CAGTCTCAT ATATATACAG TGGCTGCTTT TCTCTATTC TCTTCTTTT ATTCATATTC 960
AGCGCCAAAG AAGCAAGAC CCGTGGAAA GCTATATCT TCCAGTTGCG CCGTCTTCTC 1020
TTGTGTGCTT TCTTAAAGAA CAGACTCTTC CACAGAGAG TCTACTTCCA GTCGGCCCTG 1080
AGCAGCTCTA CTCCTGAGA GAATGTCTCT TCACCGCATC CGTCGCTCG CAACTGTAG 1140
GCTATGCGAG GTACATGAT TGGCTGCGAT CCAAGGAGGA TGGGCGGGAT TGGAGAGAGC 1200
TGTGCGAGCT CTTTTCCTG TTCAGCTGCC GCTGCGAGG GAAGGCGAGA CCGGCTCTGC 1260
CAGGGGCGCT CTCGTATATC CTTTCTCTCT GAGGATTTGA AATTTTGTG TCTGTGTGAC 1320
GTATAGCGGA ATGTGCTGTC ACAGGATGT GTGATTTTTC AACATCACCG TGATCTGGA 1380

AGGACGAGG GTTTTCTGC AGCTATTTTC TAGCAATTCG CAGTCCCTGT GCTGTGACTG 1440
ATTGGAGAC TTGTTTTC TCCTGTGCG ATTTACGCTT CCACTTTTCC ATCTGTCTTT 1500
CTACACGCT TGGATGAATG GATTTTGTAA TTCTAGCTGT TGTATTTTGT GAAATTTGTA 1560
AATTTGTGTT TTTTCTGTCA AACATATCA TTGGATATG GAGGTAAAGG AATGTGCCAG 1620
TTGCTCTGCG TCATCTCTTT TATAGCCATT ACTGTCTGTT TCTGTGTAA CAGATTTAGG 1680
TTTGTGCTCT TCTTGTCCA CTGCAAAAA AAAAAA 1717

(2) INFORMATION FOR SEQ ID NO: 125:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

60 CCAAGCTGCC GGTCACTATG TATGTGAGGG GCGAGACGCC TCCCGCAAT TCTGTGAAGT
TCTTATGCTC GACTAGGGA GTAGCCCGAG GACTCTGAT CCGCGGCTTC AGGTCACTGC 120
CGGCTGAGCG GAGTGGCTGT CCGATGTTT GGTGTCTGCG TGGCGGGGAG GCTGTGTCAA 180
ACAGCTGAC AGCAATGGC AGAGATAAA TTTGTTTTTG ACTTACCTGA TTTATGAAGT 240
ATCAACATG TTGTGTGTTT TATGTGGA ACATGCCAT TTCTGTAGGG AATGTGAGA 300
TCTGTCTACT TTTCTTATCC TGATTCATAT GGAATGGCAG TATGGCACT CTTAGGATTT 360
GTCAAGATG GGAAGCCAG TCCATCTTC AATATTTGAG GTCTTATATC TGGAGAGGCA 420
AGCCAGATC CTTTGGAGC CAGTATATAT GTCCGAATC CATCTGTGCG TCAGATTGGA 480
ATTTCAGTGG AATATATGA CAGTATGCT CAGGAGACT CTGTAGGTAA TGTCTGTGTA 540
TCTCAGTTG ACTATATAC TCAATTGACA CAAGAATGT TGGCAATTT CTACATTTT 600
GCTTCATCAT TTGCTGTGTC TCAAGCCGAG ATGACAGCA GGCATCTGA AATGTCTATT 660
CGCGCAATG TGTGTCTGAA ATGTATGGA AACTTTGAA GAGGACTAGC ACAGAGCCT 720
ATTTTGTGGA AACATATAT TGAATTAAT AATTTTATAT GATTTATGTA AAAAAA 780
AAAAAAAA AAAAAA AAAA 804

(2) INFORMATION FOR SEQ ID NO: 126:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 431 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

5 GGCACAGCCC AGGCGCTTGA AGCGCTGTGG CCTGTGGAGAG GGGCTGTCTGT CCGACCTTGG 60
GGAGGCTCTG GCGTGGGGCT GCGCTGTGAT GCGGTGATCT GCGAGCATCT CCGAGCATCT 120
GCTGGGGTGA ACTTTATTTT AGCCCTTCCC TTGTGTCTCT TATGGAGAA CAGAGAGGG 180
GTGGCGAGCT CAGTGTATC AGCATGTGAG TGAATTCGAG CACAGCGGCT TCTGGGAGA 240
GGGATGGAG GCAATTTCTT CAGGAAATG GTCAATNAT TTACCCAGAA GGCATTCAT 300
TAAATTAGT CCGAGCTTT TTGCGCCGAG CTCTGTGTTA TTAGGGGCC TTGGCGAGA 360
CTTCAGGAG GGGCGAAMN GACCTTTAAG TTTTAACTTT TAAACAGAG AACCCGAAA 420
GGTTATTTT G 431

(2) INFORMATION FOR SEQ ID NO: 127:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3752 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

35 NCGACAGGG AAGTCACTT GACTAGAA CTAGAGATAT CCAATGACC AGAAMATTT 60
AACTTCAGC TTCTTAGCA TAAGGATTT CAGAGACTC TTGGTGGCA GAGCCTGTG 120
TATGATACCA GATTTAGAAC TGGCAGACA CTGAAGAAA AGACTTTTCT TCCGAGAT 180
ASTCAGAAC TTGACATTTT CTTAGGAAA GTACAGACA AATGGGATAC TGTTTTGGC 240
AAGTCTGTGG AGCGGAGCA CAGTTTGGAG GAAGCCCTGC TCTTTTGGG TCAATTCATG 300
GATGCTTTGC AGGCATCTT TCACTGGTTA TACAGCTGG ACCCAGACT GGTGTAGAC 360
CAGCCCTTGC AGCGGGACC TTGACTCTGT CATGACCTC ATGGATGCAC AAGAGTTT 420
CCAGAGAAA CTGGGAAAG GCAACAGAA CCGTTCAAGT CCTGAGCGG TCGAGCGGAG 480
AGCTGATTA GATTAATCCA GATACACCA CTTGGGTAAA AGCAGACTC CAGGAATGA 540
GCACTGCTG GGAATCTC TGTAACTCT CTTTTCOA ACAAAGCCG CTTGAGCAG 600
CCTTAACAA AGCGGAATG TTTCAGACA CATTCCANT GCTGTTGGAG TGGCTTTCTG 660
AAGCAGCA AACCTTGC TTTCGGGAG CATTCTCTG ATGACAGAG GCGCCTGAG 720
TCTCTCATG ACACCCATA GGAATCATG AAGAAAGAG AAGAAAGAG AATGACCTT 780

5 AACTCAGAG TAGCCATGG AGAATCATC CTGGCTGTCT CCGACCCGGA TTGATACAA 840
ACCATTAAC ACTGATCAC CATCATCCGA GCTGCTTGG AGAGGTCTCT CAGATGGGCT 900
AAGCAGCAC AGCAGGCTCT TGAAGCGCC TTGTGAGAC TGTGTGCTAA TGTGTAGCTC 960
CTGAGAAC TTGTGCATG GATTCAGTGG CATTAGACCA CCGTATTTCA GCGGATCAG 1020
GAGCTATTC CCGAGACAT TACCCAGTT AAGACCTTTA TCGTGTAGCA TCAAGCATTT 1080
ATGAGGAGA TGAATTCOA ACAGCTGAC GTGAGCCGG TCACCAAGAC ATACAAAGG 1140
AAGAAATAG AGCTTACTCA CCGCCTTTTC ATAGAGAAAT CCGCAGCGG AGGAGGAAA 1200
TCCCTAATG AGCCAAAGCC TCTTCCATG CCAATCTTTT CAGATCTGTA AGCAAAAAC 1260
CCAGGATCA ACCAGCTTTC TCCCGCTGG CAGCAGATGT GCTGTTAGC ACTGAGAGGG 1320
CAAGGAAC TGAATGATG CTTGATGCG CTGGAGAGT TGAAGAAT TCCCACTTT 1380
GACTTTCATG TCTGAGGAA AAGTATATG CTTGTGATGA ATCAGAAA GTCTCGAGTG 1440
ATGATTTCT TCCGGGCAAT TGAATAGGAC CAGGATGGGA AGATACAGG TCAGAGATTT 1500
ATGATGGCA TTTTATGATC CAGTTTCCC ACCCAAGAT TGAAGATGAC TGTGTGTGCT 1560
GACTTTTGG ACCGAGATGG GATGATTTAC ATTGATTTT ATGATTTT GGTCTCTCTT 1620
CATTCAGCA AGATGCTTA TGGACAGCA ACCGATCGAG ATAAATGCA AGATGAGTT 1680
ACAGAGAG TGGTCTATG CAATATGTA AAGGTTTTC AGTGTAGCA GATCGAGAG 1740
ATAATATCC GGTTCCTTCT CCGCATCAG TTTTGGGAT TCTAGCATTT GCGGCTGTC 1800
GTATTTCTG CCAACCTTGA TGTTCGCT TGTGTAGGA TGGATGGCT TGGATGATTT 1860
TTTATGAAA AATGATCCT CCGAGCAGG AGGTAGAACT AAGATTGAC TTAGAGAGA 1920
GTCCAAACA TCTTCCGGG CAGCTTCCC TACTGTTTC AGTCTCATG GGGTTCGAG 1980
ATTCATCTTA CAGAGGGAG CATCCAGGG AATGACCCC TTCTGCTCAC GGGTTCGAG 2040
TACCGAGG TGTATCTTA TGCATCTTC TCCAGCCACC CCAAGCATG GAACCAAGT 2100
TATCCATCA TCAAGTATGA AATTGAAAG ACCACAGCA ACTTTTCTT CTATCTGAC 2160
ATCCCTTCT GGTATACCA GCAATGATTT CTTTCCCGGC CTCACAGT GCCAAACTA 2220
ATCGGGAGA CCTTAAGAG TGTCCAGTC CCGCTGGAG TCGGCTGGG AGTTCAGCG 2280
GGATGTGAC CAGAGCGGG CAGAGATG AGCTTTCTGA CTTTGAAGCT TTAGAGAGC 2340
ATTGCTTTT CCGACTTTC AAGAGCAGC GTTGCAGGG CCGAGGCAA CTCAGAGAGA 2400
GGGTAAACA AACCTTCCA AATCCAGC ATGCTTAAGA AGACACAC TCCCTCCGC 2460
AGGATCGAG GTCCAGCG ATAGCATGT CTAGCAGCC CAGCCACT ATGCATTTG 2520
AATCTGCTC CATATCTG GTGTATATT ATTCTGAAG GAGAGATTA TATGTTAAA 2580

5 AGCTGAAAG AATGATTTG TTATGAGCT GCTTATTTT GTAGTTACT 2640
ATTTCATCT GATATTTAT GTAGATAAA TTTCCTCTCT GTTAACTCTG TATTCAGATG 2700
GCTCCAGAAA TGAAATATTT GAGAAAGAA AGTGAAAGG TCAAGTACA AATGTATAT 2760
AAAAAAAAA AGGCTATTA ATAGGTTTC TCGCGGTGC AGGGTGTAA ACTGCTTTA 2820
TCTTTTAGA TTATCTTAA ATGCATCTC TTATTAAGT TGACTGTCTA TCTCAGAG 2880
ATTAATATA TTAAATAAAT AGCATCTCT GAGTGTTTA GGAATCTTT TTTTGTAAT 2940
CAGCGACCC TCATATAGA AGAATCAGG GAGGCGCTT TTCCATTTG TATTTTTCG 3000
TGATTTTAT CTAAAGAGG GGAATCTAT CTATTAACA TTTCAGATG ATACAGAAA 3060
AGAGTTTCAT TGCATATCT TCTTTGATA TTGTTCACT ACTGGTGTG TAAAGACAA 3120
ATAGCTGCTA GATTCAGGG GTAAATGTAA GTGTTGAAA AACGTGAAA CATTTGGGT 3180
TTTAACTGA TTGTTGCTC CCAATCAGC CTAGACACA GTAACTCTG TTTTCACCG 3240
GACCGACCC CTTCGAGGG GATGAGCTG TTGTGACAT TGTGAATTC AGATTTGTT 3300
TATCCACTT TTGTTATAT TATTAATG GTGATCAG TTCCACAAA CTGAGATG 3360
AATTCAGCA CCTGTTCTG AAATGTGGA CTATAGCAA AGAGTGTCT GTCTTTTAT 3420
GGAGTTACC AGCAGACTG TTAAACAGC CTGTTGCTT TGTCTTTTT TTGTGCTTA 3480
TAAATGAC TCAGCATG ACATGAAA GGGGTGTCT GGGGTCTCTG TTGTTTACT 3540
GCTGTTCTC AGCTCGACC ATGTTGCTT GTGATTTCT CAATGCTT TATTTGAGC 3600
AGAACTGAA CCTCTACAA TGAACTGTT AGAAGCAGA CAGACTTTG TATTAATAT 3660
GCTTCAGTA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTCAGG GGGCGCGGT 3720
ACCAATTCG CCTATATGA TCTTAACAA TC 3752

45 (2) INFORMATION FOR SEQ ID NO: 128:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1144 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55 TGAACCTCTG CCGCGCGGCG TCAATGCTG AGGCTTTCTG TTTTGTGCA GTGCTCTC 60
GTTAAACAA GCGGCTGTG GTCAACACT CATTACGA GCTTCATTTG GTTCAGAAC 120
CTTAGACCG CTTGCTCTC AACTATCAG CAACAGCT GCAACATC GGCCTCATG 180

5 GACAGAAA CTCCCTGCT CTCAAGCTC CTTCAAGCTC CAGTCCAGCT GACGACTTGG 240
GACAGACTA CAACCTGTG CGATATGGA GCTCCAGAT TGAAGAGAA AGCTCGAGCC 300
CTTGTCTTAA TTCCACTTT CTCAAGAGA ATTAATTTA GCAAAAACA AACAAACATA 360
GTGCGCTCT GTTAGATCA TATGTGCGA GTTTCGAGA CATCTTTTAA AGGCTCTTAC 420
TCCAGCTCC CTCCAGCC TCCCTCTCT TCCAAAGAG ACCGAGCAG GGCAGCTCA 480
GACGACTGC TTCTTCAAA TCTTCTCTG AATTAATATA AATTAATAT TCTCTTTG 540
CTTTTAGAA AAGTCTGGA CTCAAGACA AACTCTATA AAGCTGTAC ATCTAGAAC 600
CTTCCCTTT ACTGCTTT CAGCACTCT CTCCGCAAG CTTTATTTAT CTGTATGAC 660
ACAGATTTGA CATTAAGCT AAGCAATTA TTGAGTTGA TTCAAGATC CTGCAATG 720
ACATTTTCT TAAATFACA AGTTGCTTT TTAATATTT CTCATATTA TCGCCAGAA 780
TCTAATTTTA AAGCTGTAT AGCACTTTCT GCTAATAATA GATATATTT TTAATATAG 840
GCTGTCTCTG TTTAAAGCA GATTAAGAA ATTAATGTA ACTTAAGAC TGTAAATGA 900
TGTAAAGCA TTCAATGAG ACATATGCA TTTTCTCTG TTTTCTTACT TGAATATAT 960
ACATTTTGA TACTGACT TATTTTAAG ATGACTGAA ATGCAATAG CCAATCTCTG 1020
AGATACAGA TTGAATGCT ATTTCTTAA AATACACTT TGTGTTGAT TTTGAAATTA 1080
ATGATGCTTT TTTCAAAA AAAAAAAAAA AAAAAAAAAA TCGAGGCGGG GCGCGGTACC 1140
CAAT 1144

(2) INFORMATION FOR SEQ ID NO: 129:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

50 GATTCAGAG GAGCACTCT AGGTTTCTC TGAAGCAGC GGCATTTTG GCACGAGCC 60
ACGGTTTCT GCGCAATGC ATCATCAGG CCAAGGAGT CATTCGAGAC AGGCGCCAG 120
GTGAGATTT CTGCGAAA ACTTCAGT CAAGACAG ATGCGCCAG GTGCTCGCG 180
CTCTACCTC CTCTGCGAG CGCGGTGGA TGGGAGCTC TGAATGCGA AATGCGCCAC 240
GCTGCGCGCC GGCCTCTGGA GCTGGATTT GGAAGAGAC AGCAAGCAG GCTGCGCTTC 300
TCCAGGATG GCGCAAGCT TCGCAATCG CCGTTTCGG GACTTCCCA GCTGCTCTCC 360
TGTCTCTTC CCGGAGAGC CTGCGAGCC TCGCTGTAT GACGAGCTGG CTGATTCGCC 420

CTGGGCGGCG CAGTCTTCTCA CAGCCCTGCC AGAAGCTGGA GGGGTCTCTGG AGACCCATAG 480
AGCTGATGGG AGCAGCTGGT GCTTGGCTTT GGGTCTCTGG GTTCCGAGAA CCGAAGGGAA 540
GTTTATGAGG GCGAATGAGG GCGACCGGGC TCCCTCTGGGA TGGCTCCGCT GCACTTTTGA 600
AAGCCCGGTT TCCCTTCAAG TCCCAATGCC AGGTGACGAC AGCTGTCTCC TCCCTCTCAT 660
CTTAGCTTCC AGGTTCACCC TAAACCTGTA CTAACTGCT TGGTGGACTT GGAATAGACT 720
TGGCTCTCTG GGAAGAGAG AGAGCGGGCC TCCATCAGCC CTGTTACGAG AGGATCCCGG 780
AGAAGCAGAC CAGCTCTGGA CATCAGCGCC CCTGGAGCTG GGGCCAGCAG CCGTGGGZAC 840
GAGATTGGCT CTGACTTTAT TTATATGCGA TGAATCTCT GATTATTTT GGGATTTTTF 900
GTTTCTGAGG TTGTCAAGGT TTGTTTTTC TAAAGTTTGG TGAATATATA TTTGACATTT 960
TACATTTCAA AGAAGGTAT GTTCTCTAAC AGGGAGCGAA CAGAGCTAG TATTGACMAC 1020
TGTCTCTGCT TCTACTTAAA AAAAAAGAGC ACAAAGGAAA AACTAAATTA TTGAAATTT 1080
AAAAATGTC ATTCTTCTCT GTTCTTATAT ATTAGGGTTG TTAGGTCTCG TTTTGAAGTA 1140
TGGACTGTGA TTCTTTTCCC CAGCCCTCAT TCTTCAGCGG TTGGCCGGTG TTAGAGCTCG 1200
CTCTCTTTGA GTGACTGGCT AGAAGGGCTT GAGAGGTGGC CAGCCAGGCT TGGAGCTGGA 1260
GGGAGTGGAG CCCCAGCTGA GTTCCCTGTT CACAGGGTT AGAAGGTTAC TGGGAACAC 1320
CGGCGCGTGG CTCTCTGTAT TTATTTTCTT GATGTAACT TCTCAGAGCA GGGCATTTGG 1380
GACATACCA GCGAGAGAC AGAAGGACAC CTTGCTCTCT GGGAGGGGAG GACCAACAA 1440
AGCCCGCTCG GCAATTTTGT CCCCAGCTTT GGTATGCTT TCAAGGAAAG GTACAGCTG 1500
GGGAGGAGCG GGGGGAGCGC CTGTCAACCC TGGCAGGTGG TGAATTCAGG TGGGGGCTCC 1560
CTGCTACCCC CAGGCTTGGG AGCTTGAAGC CTTCCCGGCA TCTGGCATCC GAGCCTCCG 1620
CCCTCCAGGG TGGCTTTCCC TCTCTTGGCG CAGCATACAC GAGGGGAGGC AGTGGGCTTG 1680
TCACTGTATC TTGCATCAGA GACAAAGGAG GAGCCGCTTT AGCCCTCTCT GGGGAATGG 1740
GGGATGGCCC AGGGGAGCG CATTTGACAC TGGTTTAACT TAAATATGAC AGATTCTCT 1800
GTTAAATTC TTGATAGAT TTTTATAT 1830

(2) INFORMATION FOR SEQ ID NO: 130:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1864 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

GGCGCGCGG ATGGGAGCC CAGCCTGGC CCGAGACACA GGGGTCTCTGG TGGAGACTTT 60
5 TGTAGTTTAT AAGCTGAGCG AGAAGGTTTA TGTCTGTGGA GCTGGCGCCG GGAAGGGCCC 120
AGCACTGAC CCGCTGCAC AGCCATGGG GCGAGCTGGA GATGAGTTGG AGACCGCTTT 180
CGGGGGACAC TTCTCTGATC TGGGGGCTCA GCTGATGTTG ACCCAGGCT CAGCGCCACA 240
ACCTTTCACC CAGGTCTCCG ATCAACTTTT TCAAGGGGGC CCGACTGGG GCGGCTTTGT 300
AGCTTCTTAT GTCTTTGGGG CTGCATGTTG TCTGAGAGT GTCAACAGG AGATGGAACC 360
15 ACTGGTGGGA CAGTGGCAGG AGTGGATGTT GGCCTACTG GAGAGCGGGC TGGCTGACTG 420
GATTCACAGC AGTGGGGGCT GATTATCCCA GATCACTGAA GCTGAGATGG CTGATCAAGT 480
AATTTGCACT GAATTTTAAA GCGACTGTGA CTCTCTGCA AGTTCCCGAG ATCTTGAGGA 540
20 CCTGGAGCT ATCAAGCTC GAGTCAAGGA GATGAGGAA GAGCTGAGA AGCTTAAGCA 600
GCTACAGAC GAGTGAAGGA ACCAGATGAA TATGCTGTCA CTTCCAGGCA ATCTTGCGCC 660
25 GGTGATGATG TCAATTCAGG AGAAGATGGA GGTGATGTC GATTCCATCT ATGTTGGCAA 720
TGTGACTTAT GTTCCAGCAG CAGAGAGCT GAGAGCTCAC TTCTATGCT GTGGTTCAAT 780
30 CAACCTGTT ACATACTGTT GTCAAAAT TTAGTGGCAT CCCAAGGCT TTTGCTATAT 840
AGATTTCTCA GACAAAGCT CAGTGAAGAC TTCTTTGGCC TTAGATGAGT CCTATTTAG 900
AGGAAGGAA ATCAAGGTGA TCCCAAGG ACCCAAGCA CAGGCACTA CACAAACAGA 960
35 CCGGGTTTT CAGCGAGCCC GCTACGGGC CCGGACCAAC AACTACAGCA GCTTCCCGCTC 1020
TCAATTTCTAC AGTGGTTTTA ACAGAGGCC CCGGGGTGCG GTCTACAGGG GCGGGGCTAG 1080
AGCGACATCA TGGTATTTCC CTACTTAAA AAGTCTGTA TTAGAGGAG AGAGAGGAAA 1140
40 AAGAGGAAA AGAGGAAA AAAAAAGAT TAAAAAATA AAAAAAATA ACAGAGATG 1200
ACCTTGATGG AAAAAAATA TTTTAAAAA AAGATATATA CTGTGGAGG GGGGAGATC 1260
45 CATTAATCAA CTGCTGAGGA GGGAGCTGCT TTGGGAGATA GGGAGAGGCC CAGGGGATGG 1320
GCGAGGGGCG TGGTATTTCA CTCTGGGAT TGGCATGGA CAGCTCTCAA CTGCGCACT 1380
50 GCTTGGCAT GTTTCGCTCG CCGACCGCAC CCGCTCTCTC GCGCTCCCTG CCGCTCCAGA 1440
TTGCTGAGTG ATCTATTTTG TTCTTTTGG TGGTTCTTTT TCTGTTTGA GTCTCTTCT 1500
TTGAGGTTT CTATAGCGCG AGATCTCGG TTCCGCTCCC AGCGCTCCA GTGTAAATTC 1560
55 CCGTTCCGCC CACTACTCTG TTTTGGGGGG TTTAGGGGTT TTTTGTGTTT 1620
TCAATGTTT TGTTTTTTGG TTTTATTTT TTTCTTTTCT CTTTCTTCC TTTTATTTGG 1680
AGGGAATGGG AGGAGTGGG AGAGGGAGG TGGGAGTGG ATTTGTTTGA TTTTATTAGC 1740
60

5 TCAATTCAG GGTGGGAT TTTTATTAA TATGTCTCAT CATTAAGTT GTTTTGA 1800
AAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1860
AAA 1864

10 (2) INFORMATION FOR SEQ ID NO: 131:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2041 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

20 GCGAGAGCG CCGGCGAGG CCGTGGACC GCGCGCTCC CGGGATGCT GAGCAGCGG 60
CTGCTGGCC TGTGTCTGC CPTCAACCC AGGAGATGA AGCTGCTGT GCGCAGTCC 120
TTCTGGCTT AGCTGCCCT TGTTTGGGC ACTTCTGTTA ATATGAGTC TATCAGAAA 180
AATGTTGAC TAAATTTGA AAGCAGATT GAGGATGAG TTGAAACCACT AAGAGAGAA 240
ATCAGAGAT TAGAATAAG CTTTACCGG AATATCCAC CAGTAAAGTT TTTATCGAA 300
AAGATCGGA AAGAAATTT GATACAGGA GCGCGAGGT TGTGGCTC CCATCTACT 360
GCAAACTCA TGTGAGCG CACAGAGTG ACCGTGTGAG ACATTTCTT CAGCGGAGG 420
AAGAGAGCG TGGAGCACT GATGGACAT GAGAACTTG AGTTGATTAA CCAGGAGTG 480
TGGAGCTCT CTACATGAG GTTGACAGA TTTACGATCT GGCATCTCA GCTTCCCTC 540
CAAACTACAT GTATATCTT ATCAGACAT TAAAGACCA TACGATTTGG ACNTTAAACA 600
TGTGGGGCT GCGAAGCA GTGCTGGCC GTCTGCTCT GGCCTCCACA TCGAGGTGT 660
ATCGAGATC TGAATGCC CTTCAAGTG AGGATTAAGT GGGCCACGTG AATCCAAATG 720
GACCTGGGC CTGCTAGAT GAGGCCAAC GTTTTGGGA GACCATGTG TATGCTTACA 780
TGAGCAGGA AGCGTGGAA GTGCGATGG CAGAAATCTT CAACACCTTT GGGCCAGCA 840
TGCATATGA CATTGGGGA GTATGACGA ACTTATCTT GCGGGCTCT CAGGGGGAGC 900
CACTACAGT ATACGATCC GGTCTCGA CAGGGGCTT CAGTACGTG AGCGATCTAG 960
TGATGGCTT CTTGCTCTC ATGACAGCA AGTTCAGAG CCGGCTCAC CTGGGAGACC 1020
CAGAGAGCA CACAATCTA GAATTTGCTC AGTTATTTAA AATCTTGT GTATGCGGAA 1080
GTGAATTTCA GTTTCTTCC GAGGCCAGG ATGACCCACA GAAGAGAAA CCGAGCATCA 1140
AAGAGGAAA GCTGATCTG GGTGGGAGC CCGTGGTCC GCTGGAGAA GGTTTTAAACA 1200
AGCATTTCA CTACTTCTT AAGAACTG AGTACAGGC AATTAATCAG TACATCCCA 1260

5 AACCAAGCC TCCAGATTA AAGAAAGAC GAGCTGCCA CAGCTGACT CCTCACTTT 1320
AGGACAGAG ACTACCATG TACACTTAT GGAATATAT TTGCTTTT TTTTGTGTC 1380
GTTTAAAGA AGACTTTAC AGGTGTCTG AAGAACAAAC TGAATTTCA TTCTGAAGCT 1440
TCTTTAATG AATGATGAT CCTTAAAGC TCCCTCAMA AAGCTGAGA TTTTGGCTTG 1500
CACTTTTGA ATCTCTCTT TTATGTAAA TAGGTAGAT GCACTCTGC GTATTTTCA 1560
GTTTTTTAT CTGCTGTGA GACATATCT TGTACTCTG GTTCAAGTT TTATTTACTG 1620
GTTTCTTGT GAGCTGAAA AGAACATTA AGCGGAGCA AATATGCCA TTTTATTTAT 1680
AAAAATGCT ACTTAATAA TGAATCTTA TACTATGAT AAGAAAT CTTAGCAGTA 1740
TTTCTAGTG GTGTGGGCC GGAATGAT TTAGGAGGA TAAAGAAAT CTGTGTAGA 1800
GCTTTATCT TCTCTTTAA TTCAAGTTT TTCCAGTCT TACTTTTGG TTCCAACTT 1860
GACTTGAAG TATCTCTGT GTCTATGATC AAGGATTT GAAATGACTA CTGTGTTTG 1920
CTGCTATCT GGGCGCGGG CAGTTGGGG GGCAGAAAT TACATATTC TTGTTTACC 1980
ATGTTAAT ATCTATTTT AATAAATAT TGAATCTAC CAATAAAAA AAAAAAAAA 2040
A 2041

(2) INFORMATION FOR SEQ ID NO: 132:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2012 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

60 TACCAAGCTC CAGAAATCTA CTATATCATG GCGAAGGAG TAGATTTGGA CTATTTGGCT 120
GACCGAGCT GCGAAGCGGA ATGCGCAAC CAGTCTGGA AGGACATTA TGGTTACAT 180
TTCTGTAGA AAGAAATGG GCTCTGGGT TCCAGTACA AAGAACTGT ATTCAAGAA 240
TACATGATG GTACATTCAG GTTCTCTGG CCAAGAGCT GACCGAGGA AGACTTTGGA 300
ATCTTGGCT CACTTATCA AGGTGAAGT GGTGATATC TGAATGTGT ATTCAAGAT 360
AATGCGGCC GCGCTTACT TGTGATCT CATGAGTGC TAGAATTTAC TACTGTCTGG 420
GCACTGGCT CTAGAGCTGG TGAAGTGTG ACTTATCAT GGAACATGCC AGAAGATCT 480
GGCCTTGGC CAAATGACT GTTGTGTGT TCTGTATTA TTATTTCTGA GTGATTTCCA 540
TGAAGCAT GTATATGTC CTGTGGGCG CTTTGGCTAT CTGCCAAG GCAATCTGG

5 NAGGCCCATG GAGGACGAG TCAATGAGT CCGGAATTTG CATTTGTTT CTTCATTTT 600
GATCAAAATA AGCTTTGTA TTTCAGGAA ATGTGGCA CCATGGGTC CAGGATCCA 660
GCCATATTA ACCTACGGA TGAATTTTC TTGAGAGCA ATAAATGCA TCCATCAAT 720
GGAAACTT ATCCCACT TGGGGTCTT ACCATGACC AGGAGAGG AGTGGGCTG 780
TACATCTGG CCATGGCCA AGATGAGT CTACACCA TCCACTTTA TCCAGAGAC 840
TTCTCTATC GGATGGCA GACTACGG GCAGATGAG TGGATCTGT CCGAGGACT 900
TTTGAGTTG TCGATGAT GCGAGGAC CTTGGACAT GCGTATGCA CTCCATGTG 960
15 ACTGACATG TCCATCTGG CATGAGAC CTCTTCACT TTTTCTCTG ACAGAGAC 1020
TTAGCCCTC TCACTCTAT CAGGAGAG ACTGAAGAG CAGTGGCCC CAGAGACAT 1080
GAGAGGCA ATGTGAGAT GCTGGGAT CAGATGCCA TAAAGATGT TCGATGCTG 1140
GCTCTGTT TGTGTGCA TGTGTGAC CTCTGCTG TTGTCTGCG TGTGTGTGA 1200
GTCTTTGCT ACCACATG ACAGAGAG CTAGAGGCA ATAGAGGTC CATCTGAT 1260
25 GAGACTTCA AGCTTCTG TTTCAGAG TAAATCTG AGCTGAGG TATCTGAG 1320
AAGACATCT GTATGCACT CCGAGGAG CATGCACTAG TCACTAACC CAGACTCAA 1380
GCGCATGCG TGTGTGAG CAGAGAGAG CATCAAGCT TATCTGATA TTCTTTCTT 1440
TATTTATTT ACATGAAAT ATATGATTT CACTTTTCT TTAGTTTCT TGTCTAGT 1500
GCGCACTCG CACTAGGGA GTACTTAT ATCTGATC GCAATTTCA AGCTGAT 1560
35 TATATTTCT TCTGACAT GAGGTATT GAATTTCTA GAATGTATC CTCTGACA 1620
ACTAGAGCC AAGAGAAA CTCATGAT GGTTTCTAC TTCTTCAG GACTCAGGA 1680
ATTCACTTT GACTGAGC CAGTGAAT GTTAGATA CCGACTTA ACTAAGGC 1740
TAGAATATA GCTTATGCG GAATGAGG GTAGGCTG TATTGGAT CCAATGGA 1800
TTTGTATCT CTTTGGAG GACTACTT GAGGAGTGG TCAATGGTT GTTCTGCA 1860
45 TCGCATGTA CAGCTCTG AGCTAGAG TCTCAGGA AGCGATTT CCAATTTCT 1920
NAGCTGCG ACTGAGGA ATGTGATT ACCTTTAT GTTTTAGA CCAAGCTTA 1980
TCCATTAAG TACTGTTAG AAGAGAAA 2012

(2) INFORMATION FOR SEQ ID NO: 133:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1669 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

5 GAGCATATT TTACCACT TGTATTACG ATGTACAGT TCAATTTAG AGTCAGAAA 60
AGACTTTT TGTCTTTT CTCTGATGT GATCATGTT TTGTGGGTC TTCTATGCA 120
CATTTACTG TTCTGCTGC CAGATGTTGA GCGCAGTCT AGCTGAGAC ATCTAAGCG 180
AGGACAGCG TGTCTGCTT TTCTTCACT TCCCTGCCC ATATAGCAC TCTCCAGGT 240
TTAGATTACC GTTTTCGAG ACAGATTAC CAAAATGCC CAGACAGGT TTTATTACT 300
TTATATATA TACTTTTAC AGTACAGAC CTAAATTTTA TTATTTTGT CTCCCCCAT 360
CTGATACAA ATGTTTAAAG TTGTTTAAA TCCAAATG GTAGTTTCA TGGTTAAATA 420
TTTCTAGCG TATGTAGAG TTACAGCCC ATAGCATAGA AGTAAATCAG TAGCATCTGA 480
20 GACTTTTGA GGCATAGCG CTTCTCTGG CCTAAGACC TCACTTTCCC AGCTCAGCT 540
TCTCTCTCT TCACTCTCC ATCAGGCTG TTATGCGAC CTGTATGAG CCAAGTCTG 600
GTCCAGGGA ACAGCAGCG TTAACTGTC TCCCTAGAAC TCAATAGTC AGTTTAAATC 660
25 ATGATGAAC ATGATTTAT TTATGTTTT ATATAGCTTT CTTAGACATA CCAAGCTC 720
ATTGTAAT CAGTAAAT ATTCAATTT TGTGTTTGA AGCTAAGTA TGTGTAGCT 780
30 GAAGAAAA TGAAGTGT TTCTCTCTG TTATCTAGA GTGTGAGT ACACATGCT 840
GATTAATTC ATGTTTCAG GCGCTTGGC ATCTGCCAT GACTGATTC CAGAGAGAA 900
AGCCAAAGG GAAGCCGAG ATTCTTTTG AGTAGATG GGAAGAGCC CATTCAGGA 960
35 TATGAGTCC TGTGAATTC AGTTGTGTT GTGCTCTCT GTTAGAGTC ATGTTGACT 1020
GCTTTAGGA GCTCTCCAT CCACTTTTA CATGATGAG GCAAGATGT CTTGTAGAT 1080
40 TAACTTTGG ACNATGCG TTAGCTTGA GAATATGGA GCGCTCTCT GAGCCAGCG 1140
AAGAGCCA GTACAGAG CAGAGCGTG CAGCCTCTCT TCCCTTTCT TTGAGAGCG 1200
45 TGGTGGCAG AGCTGCGCG CTAACTCTG CTGAGCATA AGTGAATTT GCTTTTGGG 1260
CTTATCTG ATNATGCTG GAGCCTTCC TCTCACTGC TAGATCGAAC CTGGAATCT 1320
TCACTACT CTTAGTGT CAGTTTCTAC GTGTGAGAG CAGCTTGTG GCGCATGTC 1380
50 CTTGTAGTG CTGTAGACT TAAAAATTA TTCCAGGTT CCGTGAANA CAGTCCGAG 1440
GCTTCTATG ATCTGTAGT TCACTTGA TTATGACTG TTTTGGGTAC CTGAATTTT 1500
55 ATTGTTAG CTAAATATA GTCTGCGTG ATCATGAGA ATCTTTCTG GTCAACAGAT 1560
CATTAATTC TATCAGGAG TTCTATCAG GCATCATGT CAGTGTGCT ATCTGTGTA 1620
CAACTGGA TTTTGAAT AAAAATTTG TCAAAAAA AAAAAAA 1669

(2) INFORMATION FOR SEQ ID NO: 134:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1565 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

5 CACTTTTGGT ATATACCTA AGTGATACC CTCCTTTAGT TACTGCGAA ACTGTGGCT 60
15 TGGTTTATAT TGCATTATAC ACGATTACAA AGCTGTAATG GTGTCTTTT TCTCTTGTG 120
ACGGATGTTG TAATCAGAG TATATACAT GTGTGCTGTT CTTGTTCTG GAGTTTCATG 180
AGCATTTACA CATGGCTTC AGTGTCTGTT ATAGATCTGC CTACCTTTGT GAATTCATCT 240
GTTAAACCTT CTTTCTTTCA GAGAGACGG CGGATGTTG TTACTCTCTT GTGTTTCTC 300
TCTCTCTAC TGGTTATCTT TGAATTAAAG AGAGACTGTT CAGCTCGGTT GCTTTATCAT 360
25 GAATATGTTG TGTGACCTTG CAGTCTTCC AGATTGAGC AATCAGTGC TAGCTTCTCT 420
GACCAAAAT TAAGAGGAA AAGACAGATT TTTAAGGAA TCCATCTTTT AAGAGCGGAA 480
ACGGATGTTT CTATGTTGCT GCACCTTCTT GTTGTACTTC TGAATACGAA CGTGTGTGAA 540
30 GATCATTTTC TGACTTAAAC GTGATATGCT CAGGATTAAC CTTCTGTTTG TTTTGTTAGC 600
ATTGAAATGG AGACATATTA TTGTGATAT ATACAGAGT GTTTTTCAC TGTATTTCTAT 660
35 TTGCAAAAT TGAGATGTC TTCTCTTACC TTTTGCAAA TAAATGATAT TGCATATTTG 720
ATTCTCAAG ACTTGATAT GGTGAGCTTA TTAACTTAG AATTTGATAT CATCTTTTCA 780
TGACTGTGCG CTGAGTTCCC CAGCCCCCTCT COTCTTTT TTATGATGAG ATTATGACAA 840
40 CTCCTCACTTA TTAAACATG CAACATTTCT TGAGTATGTA TTTTGAGGCC ATCTGAGCTC 900
ATAGCTGATT CAGTAAACAG TTTCATGCTG TGTGATTCAC ACTCACTACT TAAATCTGCC 960
45 ATGCTGAAA TGTGAGGAA AATGATATCC ATGTGTGCTT GGGAGGCTA TACACTTTGA 1020
CATTTTITAA TACTGTGATT CTGTACATAT TGTGAGTTT GTTTTGTTTT ACGAGAAA 1080
50 AAAAAAAGT GATTAAGCAA TCGAGAGCC AAGAGGTTTA CTATGATGC TTAGGCTGCT 1140
CTGACCTTGG CTGGCCATA GACTTAAGC GCCAATTTAA TTATAGAGAG TAATATTTTT 1200
TCAAAAGCCA ATTTTITTTT TGTATTTCTT GTATGAACTT GCGATATCA TGAATAGAAA 1260
55 GGGAGAGCAA TAAGAGGAA AGAGCTGAT GTTCTGTTAT GTTCATGTAA ACTTAAGAAA 1320
ACAGTGTGGA GCGAGGGGCG ATCAGCGAA CTCTAGGGAC TTGTGTGTTG TTGGAAGCCA 1380
60 TCCATACCTG CATTTTCCAT TCTTGATATG TAATCATATT GCCAAGAGCA AACTATTTCA 1440

TCATTATTTG TAATATACAC TTTTCCCGAG ACCTACGATA AAGTTTCTGT CATGTATTTG 1500
CTTCCAGTTG CATTAATAT TACTGAGTTG CATCATTTGA AGAAAAA AAAA 1560
5 CTGCA 1565

(2) INFORMATION FOR SEQ ID NO: 135:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2007 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

20 TCTAAGGCC CCTTATACC CGACTTTGTT CAGCAAGAT CCCCTGCGAG GTACAGGCT 60
GATTGTGGC CAGCTGCGAC AATTCCTGA GGCACACTT GGTTCAGTT CAGATTTCGA 120
25 GCTGTGTTG TTTTGGACC AGCAGAGGC AAGCTTCCAG CCAAGCAACA GAGCTGTAG 180
AGGACTCTCA GCTAGCTGCC CTGTGAGAC CCCAGGCTT TGTCTAGACA GTTGTGTTAG 240
CTTCCGAAA GTACAGGTTG ATTGATTTG GCGAAGACTG AATATTCACA CTTAAGTCTT 300
30 GAGCATATCC TGACTTTTAC TTCTTTATG CTTCGCTGCC AGTTCTCTC TCTCATACAC 360
ACACAGGCC TTGCTCGAAA ATACAGAGC ACTTCATGCG CTCAGCTAT GCGAAGAGCT 420
35 GCAATGTTGCG TCCCTTTCTG TTTGCTTAG GAACTGTTGT GCTTCTTTG GTCTCACTCG 480
CGAGGAGCT GCGAGGTTG GACTGCGATT GGGCTGCGAG CAGCTCTGGG ACGGCGAGG 540
GCGGGGCTC TGATCAGCTC GTTGAAGCA CACCTCTCTC TTGGCTCTCT GGCAGTTCTT 600
40 TCTGCGAATA GTCTCTTCCC TGGCAGTTG AATGCGGAAA GCTCTGCGCA CAGAGAGGAG 660
AGGCGATCCC GCTGAGGCT TAGGGAATTT GTGAGGCGCG CTCGAGCGAG ATATTTGACT 720
45 GGGGAGTTTT TCGAGCTCAA AGATCATCTT GCTGTGTTG GGGGCGAGT GTGTGACACA 780
AGCATCTCAA AGTCAGAGC CATCTGCGGC TGCCTCTCTT CTTTCTCAGG CTCTGCGAAA 840
AGGATCTCCC CTCTCTCTC ACTGATTTCC AGTGTGTTG GATTTGTTG GAGCACTGCG 900
50 ACTTTTITTC TCTTTTCTT GATGAGCAA CATGCAATAT GCATCTGCC CATTTTACTT 960
TCAGTGATAT TTCTTTCTT GATGAGCAAT CTTTGTGCGC CTTTATAGAA GGAAGAGAT 1020
ACACTACGA TGTCCGAGGC ACTGTTTAG GCGCTTTTAT ATAGTCTCT GTTAGGTTGA 1080
55 GACTAAGGGA TGAGGAGATC TCTTTATAAA AGGCGCTTAA GTAAATGATA AAGAGAAACA 1140
CTTAGAGGTT AGAAGGCTG TCTTCAAGAT CCAAGGTAG ATTGCTCTCA GTCTGATTTT 1200
60 TTTCTGAGG GACTTATCCC CTACAAATAT TCTCAGTCC ATACTCTCTC TTCTAGCCCA 1260

5
10
15
20
25

CCATGCTGCT CCGTGCAGCT CTCAGATGCT CAGAGGGGTA ACCGATGCTC TTAGAGCAAT 1320
TGGGACCA TAGAATATCT GATGTGTGAA TTCTCTTTAA AAAACTTAA GAGTCTTTGC 1380
TACCTCTGC TTGTGAGGTT GTTTGCTCTT TCAATTTAAA AGCGAGCTC TCACATTTTA 1440
TTGAGAGGTT GGGCTGTGTC TGTGCGATG TGTGTATACA TTTCGAGGG TGCCTGTGTC 1500
CTGTAGCTTT TTAAAGGAA ACCGATGCT CCGACTATGA ATCTGGCTAC TTCTATGCT 1560
TCTAGTGTTT TGGCATACA TCAACGAGG GGTTTAAATTT ATCGATGCTC TGAAGCAATG 1620
TTGAGAGGG GCTGATCAA ATTTGAGAG GGTATATGGA AGGGAGGGG GAGAGCAAT 1680
TGACATTTAT TTATTTAATTT ATTTAAATG TTATACATCTT CTTTATGTG TATCAGGCT 1740
GAATAGAAAC TCAATGCAAT AAATTAATCT GTTCTGCTCT CTTCTCTGCG TTCTCTTTT 1800
TTTTTTTTTA AATTAGCAAT AACCAATTTT TGTTCCTAAA GTGATTTGTC ATTTGTGCTG 1860
TATTAATCT ATTAAGGCTT CTGTTTAAA AGGTGCAATTT TCAATTCCTCT GGGGCAATG 1920
GTGCGAGA CATCTACAT GTAGAGAAC ACAGTGCAG ATCTGTGCTCT GATCTCGAA 1980
AATTAATTC TCTATATGAT TAAAGT 2007

30
(2) INFORMATION FOR SEQ ID NO: 136:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1291 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

40
45
50
55
60

CTTTTAAACC TCCGCTCTCA CACCAATACA TATCAGGTTG TTTTCTAGTT AAAAAGCCAA 60
GTAGCTGAGA TTCTACTTTA ATGTCAATGC AGATTGCAAT TGATCAATGC CATTTGTTT 120
TTTCTCAATTT TTATGCTGTT GGGTCTTAGT TTTTAAATG ATTTAAAGAA CTCAGCAATG 180
GTTTATTTT CTACTGATAC TTAGGGTTTA GGAACACTA CCGACTAGTTA TCTTTTATTC 240
AACTCAATG GTCTACTGAA ACAAATGTC TTACTTTTCA TTATGGCAAT ATTTAGAGTT 300
ATAGTAGTGG TTTCAGAAA AGCTTCTTC ACATGTGAC TTCCCAATCA ATATCTGCA 360
TCAATAGATT ATTCCATGCA AAGGCAAGT GTTTGTTTCA AAATTAATCT AGTTTCTGCT 420
ACATTTAAT TTGAGAGGT GAGCAATGCG TCTTTTTCAG TCTTCTCTTA TGTGAGTTT 480
CTGATAGACC ACTATTTGCA AACGATATCT GTCACTAGC AAATGTGTAA AATTTCTCT 540
ATTGATCTTT GTCTATATGG TAAATATGTA ACTAAATCTT TTGGCAATCT AGCAATCTT 600

660
720
780
840
900
960
1020
1080
1140
1200
1260
1291

GCTAGGCTGG TTTTAAAGC TATGTGTAT TCTTACTAAT GTTCTATACA AGAATGCAAT 660
TGTAAATAT GCTGTCTAAT TCTAATGTT CAAATCAAT TTTGAGGTTG TATCTTATTT 720
TATATAGAAA CAGACTCTTC AAAAATCTT CAGAGGAGCG TTATATTTTA AATATCGAAA 780
TATTTAAATA AACCGGTGG GTTAGATTAC TCATCTGTCC ACCAATGGG ACATTTGCAAT 840
GAGTGTGGGG CTTAAAGAC TTAGAGAGGA CTTGTAGTA ATCTCTGAAA ATGAGCCAAAT 900
CCCACTTTGA ATGTTACTG GAGTAAAGCC ACCTTTACCA CCCCAAATAC AGCAGCGAG 960
GCCAATAAAC CAGCTGTGCT CTGATCAAT TTTCTTTTCT TCATTTGTGA TGTCTAGAT 1020
CAAAATGCT GTTCTACAT GTTACAGGCT TCTCTTTTGT TTCAATTAAG ATTTTATGTC 1080
TACTTTGTA TGGACCAAT AGAATTTCA CAGACGAAA TAGAGCAAT TGTGTTTAGA 1140
TATTTTTCAG AAGTACGAG ATTTGTGCA ATCAATTTAT TTGCTTTTT AAAAATCTAT 1200
TTAGCAATTT CAGAGTAG AGTACTGAGA AAATATTTTC AGTAAATGTT CTAGAGGTC 1260
AATATTTTT AATCAATAT GATCAATAA A 1291

(2) INFORMATION FOR SEQ ID NO: 137:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1906 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

60
120
180
240
300
360
420
480
540
600
660
720

GGCAGAGGA CTTACTTTTG TTACAGACA TGGTTGTGTC CAGGTAAAA CCAAGTGAT 60
ATTTTGGAT GCTTGTGTCG CATCTGTGAC TTGTTTTGTC AGTATCAATTA TTCAGACTTC 120
AAATGTGAA TCTTTTAAAC ATCTGTATTA TTGTGTGTC AGAGCTGTTT ATCTTAAAT 180
GTAAATGAAAT TCAGTCTAGT TCTGTGATA AGATCAATCA GTTTTGAAG GTTACTGAT 240
TTCTCTTCC CTTCTAGTTT TTACCCAAAT ATATGAGAA GAGTAAATGT CAATCTTAAAC 300
ATTTTGTGTT AATGTTTAA TTAAAGCTCT GGGCAGTGT CAGCAATCC TACCTAGTGT 360
CATTAAGGA AAATACTTAC ATAGCTTTCT TAAATATAG CAATGCAAT ACATTTTATG 420
GAGAAATTA GTTGTGTTGC ACCGCTACT TAAATGCTTT CCAATATG TGATACAAAC 480
TTTTGAATAT GGAATCTTAC TATTTGATA GAATGTGTA TGTATATAT ACATACATAC 540
ATAGCAAT ATGTGTGCT GTTGTGTAT ATATATAT ATGCAATGTC TGAATCTTCA 600
CTACACACA TAAATCACT TTTAAATTC AGCAAGGCT AGTCTGACAC GGTGATTTTC 660
CTTTGAGGC TGATGCTGTT ATTAATGTT TATTTAGGTT TTACTCCGAG TAGCAAGGA 720

5 TTCTAAGTTA GTTGGACTTA CATGATTATT GTGATTTAAA ACTAAGAAATA AAGGCTGGAT 780
TTTCAAGAT AAATTGAAAT TGTGTTTGGT GAATATACAA CAAAATTAAT GAATCTGATG 840
TACATACAGG TTCTACAGG AAGATGCT ATATTTTACA ATTGGAGAT TTAAATACCA 900
GGCTATCCCA GAATAAGTGA CTTCATACCA TGGTACAAAT AAGTAAGGGA TGTCTCTCTGG 960
10 GTTTCCTTTT GGCATCTTCA AGATTTTTAC TTCTCAGGTT ATTAATACAA ATTAATGAT 1020
AAGTAAGCA ATAGAATTTT TAGGTTAAAA CAACAGATGG GGGTTTCTGG GAGTGTTTAA 1080
TGTCAATGGC ATTTTATGTA GATAGAGCC TTGTTCTGCG ATTGGATGAT TTCTGATATT 1140
15 TTGTTTTCAC AGTTAAATCTT CCTCCGCCAA GTTTGCTATT CAATCAACT GCTCGAATGA 1200
CAATCTAGT AGTCTGATCT ATTITTTCTGA GGAATAGATT GTGATTCGA TGGAGGTGTC 1260
20 TTCAATACCA TTACCTCTAC ACTCGAAGG AAGCAAACT CCTTTATTAG AATTACTGCA 1320
CATGTTATG GGGAAATPAG TTCTGAAGG CTGGAATGAT AAGGTGAGC AAAAGTTGGT 1380
25 CAGCTTGGCT ATGGAGTGT GGCATATATC TCTAAACATT CAAAAGACC ATGAGCTGAA 1440
CCTAAACTCT CTGGAGATC TGGACAAAG GAATATGAAA ATTGCATTT GAAATCTGAC 1500
CAGCTATATC GAGCTTCAGA GATAGATAG CCAATGGGCC AAGCCATTT CAAGTACAGA 1560
30 AATTATAGAG ACTACACTTA AATAATTTG AACTTAAAT ATTAATTTAC CACTTTTGT 1620
CTTATTAAGC ATATTTGTAA ACTCAGACT GAGCAGAGT GACTTTTACTT TCTCAAGTTT 1680
35 GATTAAGTGT TGACTGTGTC CTATTCCTC ACCCTTGCC TTCCCTTTCC TAAAGCAATA 1740
GTGCACACT TAGGTTATT TTGCTTCGGA ATTTGAATGA AAACTTAAAT GGCATGGATT 1800
TTTTTCTTTT GCAAGACACC TGTATTATCAT CTTCGTTTAA TGTAAATGTC CCTTTATGCT 1860
40 TTGGAATPAA ATTCTCTTTT GTAAAAAAA AAAAAAAAAA AAAAA 1906

(2) INFORMATION FOR SEQ ID NO: 138:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1935 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

55 TCTGAATPAA TGTATACAGA TCCCTCTGAG GATTTCTTGA TGGGCTGAGC AGCTGGCTGG 60
AGCTACTACT GACTGACAT CATTCGATG AGGCGAGCTT TCTGTACAG GATTTCTAGC 120
TCTATGTTTT ATATATATTT TCACTGTATC TTGCACCTCA CTTTACAGA GAGGAATCTA 180

240 TGAAGATTAA GCTGATGCG TCAAGTGTAC TTAGGTAAAT TGGCAATGCG ATGCTTTCCA 300
CTGAGTCTCT CAGGTACAGA AGCTACTATC TGTGCTATTT TTGTATATCT TCTTTAAATAT 360
5 GTGCTAGACT TTGCAAAATC TAGAATGGGT CCTGTGTCTY TTTTACTTTT GAAGAAATCA 360
GTTTCTGCTT CTTTTTGGAA AAGAAATCAA AGTGCATATG TTTTTFACG GAAGATPACC 420
CAATGAATG AGGTGAACAG GACTAGATTT AGGCTTCTCT GTAAACAGAA ATCAATATCA 480
10 AAACCTATAT TTCCCATCTG TTTCCTCAATG CCTGCTACTT CTTCAGATA TTCAATTTCA 540
GAGAGCAGC AGTTAAAGCC GTGATTTTGT TAGTTAGGAA CTTGGATATCA AAGCCTCTTC 600
15 CACTAATGCG CTATGCTCTCT GAGCAAGTTT TTTTTFTTTT TTTTTFTHAA AGCCTTCTCT 660
AACTTCTACT TTCTATGCTCT ACCTCAAGAA ATTGTTGTGA GCTGTGAGAT AATGCAATTT 720
TAAAGGCTCT GCGAGTATGG AAGATGCTAG TTATGATTT AAGAGTTGT TAAAGCTGTA 780
20 AAGATCTAAA ACCTACAGCT AATCACAATG CATTPACCC CACTCACTTG GACATATGTC 840
AAACTAGCC AGAGTCTCT TTTCCAATTT ACTTACAGCT TATTCATAT AAAATTTTTC 900
25 TAATGATPAA TCTATTATAT CTAAACTAAA GCTTCTCTTT TATACACT CTCTTATATC 960
TGGATPAGA TAAATGACCA CATTACTTTA ATTTCTAGCT GGTGCTCTCT GATGCTTCAAT 1020
30 TGTATGTAAG GACATTTTCT YTTTTCAGC AGCTGTGTAG GTCCAGAGCC TCTGAGAGAG 1080
GAGGCGGTTA GCAATGACCC AGCAGGCGAC TGAACTGGA AACTCAAGGT TCTTTTACT 1140
GTGCGTATGT GAGCTGCTTT TCTGTGATCG GTTTCCTTAG GCAATGTCTCT GTTCCCTCC 1200
35 TTGCTATTCG CAGCTACATA CAAGTGGCC AAGCCATGA GCTGATCTCT ATATATGATC 1260
AGTCTGTGTC CTGACTCTCA ATAGGCGCAC CCAAGCTGCG TATAGGTTTGA CAGATACATT 1320
AATTAGGCAA CTTAAATAT TGAATCTGCT GTTGTGTGTA CATATGCTTA TGGCAGAAC 1380
40 TGAATCTTAG AGTTAAATTT CATGATTTAG GTTCTCTCAG AAGGACAGAA TTATGATGAT 1440
ATATGTATAT ATGAAAGGGA GTTTATTATG GAGAACTGCG TCCCATGTT AAGAGCGGAA 1500
45 GTGCGCAAT AGGCTCTCTG CAAGCTGGTT TAAAGAGAG CCAATATGCG CTACGCTCTA 1560
GTTCAAAAC CTCAAACTG GGAAGCTGA CAGTGCAGC AGCTTCTAGT CTGTGCGCAA 1620
AAGCCAGAG CCCCAGCAA CCAACCACT GGTGCAATGC CTAGATTTCA AAGCTGAGG 1680
50 AAACCTGAT CTGATGTGTA AAGGAGGAA GAGTGAAGA AAGCCAGAG ACTGAGGAAA 1740
CAGGTATAC AGTGTCTTAC ACCATATGCG CCAATACAAA GAGCTTACCG ATTCCTTCTCT 1800
55 GCTTACTGGA TCCCTGAGCT TGCCTGCTCT TCTGACTCT CTAACTCTT CTAACTTAG TTCTTAAGAG 1860
CTTTCTCATTA CATAGCTCT CTCAAGGCC TCAATTAAT TCTCATGTGA AGTTTCAAAA 1920
60 AAAAAAAAAA AAAAA 1935

364

365

1446

(2) INFORMATION FOR SEQ ID NO: 139:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

5
10
15
20
25
30
35
40
45
50
55
60

NECCGCTTGG GCACAGTCA GATGAGGAC GTTCTGCGGG GAGGCGCTTC ACTTTCGAGA 60
GAGGACAGAC ACGATTTTCC TGTCTGGGGA GGGAGGAGTC CAGCATCTCT GATCTGCTCT 120
GGAGCTTAT TTTCCTGCTGG CAGATGCA TTCTCTCTGAG TGGAGAGAG TTCTTTCATG 180
TGGATGTGT TTTCCTCCAG CAGAGCGGCC CTCTTTTCCC AGCATTTCCC TGGCTTCCCG 240
AGGCTCCAGG CCAGCACCA GTTCTCTCTC AATGCGGAG TGGAGCAGAG CTCTTAGTTG 300
GCAGAGCTGG AGAGGGGTGA ACAAGCCCTG AGGAGGCTCC GGCCTTTGCT CCGAGATTGG 360
GGGAGGGGG TGTGGCAGG TGGCGCCGCG AGAGGCGCAG CATGTTTGAC CAGAGCCCTC 420
ATTGTGTCT GAGGAGACC TTATTCCCGAG GCTTCNAGC ATTGCTCATC GTTCCCAAC 480
TGGTAGTGT GATTTCAGG GAGAGCTCC CAAATGTGCG CAGATTTTC CCGTCTCCAG 540
GCAGGCGAG GGAATTAAG GCGAGCGAG ATGAGGGCG AGGATGTGTG CAGGTGAGGG 600
GCTCTCGCCC TGTCCGCTCT CTCTCAGCA TGTCTGCCC ACCGTGCTTC AGTTCTGCTT 660
TCCCTCTCAT CTCTGTCCTT CTCTTTGAG CTGTCTCCAT CTGAGTGTCA GACGAGCTT 720
CTCTCTCAT GACGAGCTTC CTCTGAGGCA GCGCCCTCTCC TTGTCTGAAA AAGGAGCTT 780
TGAATGTGT AGCGAGGCG TGGGAGGAAA GGTCTCAGCG GACAGGTTGG GAGATGAGG 840
TCAGCGTTTC TGGGAGCAG ATGAGGGGG CAGTGGGAG AGGCTTTGGG CAGACACCG 900
CAGGATTAAT TTGAATGTG TGAAGTGAAT CCGCGAGGTC CTGGGCTTG GCGATTTGG 960
AAGAGATCA TGTCTGAGG GGTTTAGGG ACACAGTGA CAGGGGAGA GTCTCTCATCT 1020
GCTGCGATTT TGTGGGTTGT TAGTGGCMAA CTGGAATAGG GCTTGGGGTG CTGTCTTTCA 1080
CTGACACCA AATCAGAT CCGTGTCTT GATCTCCGAG AACTTTGCT CTTCAGTGTG 1140
CTTCTCTTC CTACTTCAT CGATGAAA TTAGTTAATT TCTGATCTT TCCCTTGGCT 1200
GGTCTAGCTC CTCTCCCAAC AGCATGCCC TCCAAATGCT AGAGCTGCG GCGCTGAGC 1260
CTGTAGACAG ATGCGCTCAG AATTGGGCA TGGAGGGGG GGTGGGGGAC CCGATGATC 1320
AGCGAGGAC TCCANTGCC AGCTCTCTC CCGAAGCA TCCCGACAT CCGTTATCCC 1380
TACCCGACC CTTTGGGCT CTGTACAT TTTTAACTT GCGAAGAT GAGAGATA 1440

TTCTAA

5

(2) INFORMATION FOR SEQ ID NO: 140:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1109 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

15
20
25
30
35
40
45
50
55
60

TTTTTTTTTT TTGATATCA AATTGTCTTT CTCATATCA GAATTAAGCT AGGAAACAC 60
TAACCCAAA ACTTCTCTTA GAGCTGTTC TTGAGGCA GCATCACTTA TTGCGCATTA 120
AGACTAGTA TAAAGGACC AGCATCCCTA CTGGGTCAT GGGATTAAT TTATAGCAT 180
TCCATTTTCC TAGTCCACA TGTGAAATG GATTGTGAG ATCTTAATCT ATATCTTAC 240
CTTATATTA AAGATCMAA GATATATCT CTATGACAG ATTGAGATA CGAGATGAA 300
AGTGGGAGG ATGTCTTAT TCTAATGTGA GGTGAGGGA AATGTGATA ACATTACTGG 360
GGTAGGAGG GCAATGTCT TTAGTTGAG TTCTCATTTT TATCTTCCAG TACTGACTTG 420
TCCGAAAGC ATACTTTTTC ACTGCCAGT ACTGAATCA GAGGCTCATG GAGATATATA 480
TGTGGAGAT GATGCAATTT CATTATATG CAACATAGC TCGATTAAGA CAAGTTGTT 540
GZTTTGGAA GGGTTAAG CTTTAAGTGA ACAATCTAG CTATGCTGA ATGACTTAGG 600
TATATTAAT TCAATTTTT TATTTTCTT TGTTTAAGG TCCCCATAC TTCTCTGTC 660
GGAGATGTA GAATATGAT TACTTCACTG TTAGTTTCTT TATTTTTTTT TTTCCTCTAT 720
TTCTCCCTTG TACTTTGTT GGAAGCTAGA AATCTGTGGG TTATATACATG GCGAGCTCTT 780
TGTGAAGTG GTTATATCCA CTGGGAAG GCGATTGAA ATCATTTAGA ACCATCTAT 840
TTCTTCCCC AGGAGACT ATTCCTATTA GATAGCTGA AGAGCTGCT GTGAGAGCT 900
CAGCTCCAA CAGAGATCA GCAGCTCTTA TAGGATTC CATGATTAAT GACTTCTCAT 960
TCTGTTTTAT CAGAGTGAT ATATGTCTTA CTGAGGAA AATTAAACAG TCAATTAGCA 1020
AAGAACTCA ATCTATATC TAACCATTTT AATAAAGT TAAACAAAA AATTAAAGG 1080
GACACTGAG GGGGGGCGG AATCCCAT 1109

(2) INFORMATION FOR SEQ ID NO: 141:

(1) SEQUENCE CHARACTERISTICS:

366

367

(A) LENGTH: 497 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

5 TAGGACTTAC TTAATTCCTT TTAATCATCT TTAATTTAT TAAATATTTT ATTCTTTCA 60
10 ATTTTCTCTT AATTTCTTA TCTCTCTTA TAAATGTGTA TATTCATGTG AACCATCT 120
CATTAATCTT AACATTTACT CTCAAAAGC TTATTATTTT TATTTTTTTG AAGTACTTT 180
15 TTCTGTCTCT ACTCTGTAC ATGATTTTC TTTCATATCA TTCTGTGTCC CCAATAGAAA 240
ATGCTCTTTA TTTTGTAGCA TCTGTGACTT TTTAGTATGG CATGAGTGTG CTAAAGGCCA 300
GATATCTTTT CACATTCACT GTGTGCTTTG AACCTAGTTT TTTATCTTCC CATCTTTACT 360
20 TTAATCCCTG AAGTGTGACT CTTCACTCAG GSCCAAGACC GGCCTGAGGC CTTTGTGGA 420
GATCTGAC CACACAGAGA AGGCTGAC CTTGTTACTT GTACTCTTTC ACTGTATATA 480
25 AATAGATTTA TCTAATA 497

(2) INFORMATION FOR SEQ ID NO: 142:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

30 ATGAGGGAGA GCGAGGTGC CTGCAACC CCTGCTCA GGAATGCTT TCCCAAGAA 60
40 TCCCAACCA CATAACCTC TTCTTTTCTT CTAGTCAAC TCTTCTTTAT TCTTGTGCTT 120
GCTCTCTCC TTCTCTCC TCTCAACTT TTACTCTG TTTCTATTTT ATGGGATTTG 180
45 GGTTTGAGT TAACTTACA ACGTGTCC CACACGAG TCTTCTGAGA AAAAAATCA 240
AAGATTTA AAAAAAAA 269

(2) INFORMATION FOR SEQ ID NO: 143:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1269 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

60 AATAGCAAA CTATAGGATA ACACAGACC CTTTGTGAAA ATTAATTTGC ATTGAGTGT 60

TTATATGACT ATGCTCTCTC CCGCTACAG GAGAGTCTG CCAATGTGAA GGCATGTAC 60
TTATCATCTT CCAAGACCAT CTGTGCTCTG CTGTGTGCTT GCTTGGGCTT CTTCGGCTC 120
TTCTGGCTGC TCCATGTGCTT GCGCGGAGAG GCTTACTTGC GAAATGCTGT GGTGTGTATC 180
ACAGGCGCCA CCTCAAGGCT GSCAAGAA TGTGTCAAG TCTTCTATGC TCGCGGTGCT 240
10 AACTGTGTC TCTGTGCTCC GATGTGTGG GCTCTAGAG AGCTCATGAG AAGACTGACC 300
GCTTCTCATG CCACCAAGCT GCGACACAC AAGCTTTACT TGTGTACTT CCACTGTACA 360
GACTCTGCG CCAATATTTG AGCACAGCT GAGATCTCTC AATGCTTTTG CTATGTGAC 420
15 ATACTGTCTA ACATCTCTG GATCAGCTAC CTTGTATACA TCTGTGACAC CACAGTGTAT 480
GTGTACAGAA GGTGTATGGA CACAACTTAC TTTGTGCTCAG TTGCTCTTAC GAAGCACTC 540
20 CTGCTCTCCA TGTATCAGAG GAGGCAAGGC CAGATTTCTG CCAATCAGAG CATCCAGGGC 600
AAGATGACA TTCTTTTTC ATCAGCATAT CGAGCTTCCA AGCAGGCAAC CCAAGCTTTT 660
TTTGTCTCTC TCGTGTGCGA GATGTGACAG TATGTAAATG AAGTGTACCT CATCAGGCCC 720
GGCTACATCC AACAACCT CTCTGTAAAT GCAATCAGG CCAATGTATC TAGGTATGGA 780
GTATGTGACA CCACACAGC CCAAGGCGCA AGCGCTGTGG AAGTGTGCCA GATGTCTTCT 840
30 GCTCTGTGG GGAAGAGAA GAAATGTGT ATCTGTGCTG ACTTACTTGC TTCTTGTGCT 900
GTTTATCTTC GACTCTGC TCTGTGCTCT TTCTCAGGC TCTGTGCTCC AGGCGCGAAA 960
AAGAGCGGAA ATCAGAGAC TCTTGTACT CTGACAGGC AGGCGCGAGG CAGAGAGCA 1020
35 GCACTCTTAG GCTTGTCTAC TCTACAGGG ACGTTGTAT TTGTGTAGAC TTTATGTGAG 1080
ATTTGTCTCA CAGTGTGGA AACTGTAGA AACAATCT GTGTAGATCT GCTGTGAGAG 1140
40 GACAATCAA AAGCAGACA AGCTTCTTCC CAGGTGTAG GCAATCACTT AAGCATTAA 1200
TATGTGAGCT GGTTTTACA CTAAATCTA GAAATATACA TCTCAACAG TAAAAAATA 1260
45 AAAAAAAC 1269

(2) INFORMATION FOR SEQ ID NO: 144:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1946 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

5 TTATACCTCT AGCTGTGTTTA CTATAGATGT AACATATGCT GCCTACGAC CTCAMATGT 120
CTGTACTGCA AGAGGCGCTT GCGCTCTGCG TTTCCTATAT CATGTTTGGC CAGATTTGTA 180
GTCCGAAAGA AAGAGATGCG TGGCAGATG TACGAGATG AACGCGCTG TCGATGCGG 240
ATCGAGGACA AGGGGTGACA GATGCTGCG TGCCTTACCG TGGCAGTAGG GAGACAGTCC 300
AAGATATGT CTTCGTGCA CCGGGTGTG TTGTCTCTCG TGGTCTGCA TGTCTGTGG 360
TCAGCTTTAT TCTTGAACT GAGTGTACC TGGATCTGCG TACTGAGGCT AGAGCCACA 420
GAGATATGG GTTGGCGCTG TGGGCCCCA ACTAGGGGGT GTGGTTTAT CAGATGTTG 480
CCTTTGTCT CTATAGATA GGAATCTACT TTTCAGGGA ATTGTCTC CCATATAT 540
TTGCTTTACC TTGTCTGTT CTTTGTGCG ATATATCAG TGTATATGT CTGACAGGG 600
TCATATGCG CCMAAGTGA CACTGTCTG CTGATCTC CTTTGGCAA CATCAGGTC 660
AATATCAG ATAGGCTTC CTAGGCGACT GACTTTGCG GATGCGGGG TGTGTGTGA 720
CAATATAT TCAATTTACC TGGAGATGT CAGAGGCTG CTCTGAGGCT GAGTGTGTT 780
CGCGCTTGG TGTGTCAGG TGTGAGGG ATACATCTT AGGATCTGCG AATCCAGGC 840
CAGAGACT CTGTTGTTT GTTCCAGAT CCTGTACTTA GAGGCTCTG GCGAGAGAA 900
CAGACAGCT TTTCAGAT GAGCGCAG GAGAGATG GTTTGCTGCG TGTCTACT 960
GTCTGAGC TGGTCTGTG GGCATTTGA AAGTTAGAT TGTATCTCT GGGTTTTGG 1020
TGGCTTTT CAATGCTTC ACTTAGGCG AGGAGAGA GTCTATACT TCCGAGGCT 1080
GCTTGAGCT CAATGAGG TATATGAG ATCTGTGAT ATTGTATAG AATTTCTTT 1140
GTATTTATG ATTTGAGCT CAGTCCGCA ATTTCCAGG TACATGTAA GCTATTAAC 1200
ATCATTTCT AAGAGAGG GATAGCTGT ACTCATGGA TCATGAGTC CATGGCTGT 1260
TGCAGTTCC CTGTTGCTT CTTGAGTTT TGTCTCTGCG TGTGTCTCC CAGAGAGG 1320
AGAGACTG GGTGAGATG GAGAGAGGA TTAGAGATAT AGCATATGA CCGAGATGG 1380
AACAGTGGG AGCTACTGT GATGAGGAG ATGACTGCT GAGAGCTG GAGTCAAGT 1440
GTGAATCTG TATGTGACA GAGATATAT ATCTGCGGT CTGAGGCTT CAGCTCTCG 1500
TCATGCTTT CTTGTATGAC TGGCTGCA CAGATCTGA GTTGCAGCT TGCACCTTAA 1560
ATCAGAGCT GGGCATGTC TGGGCTGCG GGTCTGCGG TTCTGTACTT GTGTCAAGT 1620
GTGGCGCTG ATGTGAGG TGGAGGTTT TGTGTGTCT GACTGTAGC TGTAGCTTT 1680
GGCTGTGTA GAGCGCTCT CTGTGCGCT ATGTGCTTC GTTTTATAC ATCATGGA 1740
AGATGGAA GTGAGGAG AATTTTCTG CCGTACAGG GTTGGAGG AAGGAGCA 1800
GTATTTAT GATTTACCA TATATCTTG TTTTCTTCA ACCAAATGT TATTTGAGC 1860

5 ANTTCTACT GCTCAAGTT GATGTGTTA TTCAATATA ACTGTAGTT TGTGATATA 1920
AAAAAAAAAAAAAAAAA AAG 1944

10 (2) INFORMATION FOR SEQ ID NO: 145;
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1021 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 145:
TCGACCAAG CGTCCGGGTT GCGCAAGCG GAGTTCCGCG TCGAGACCG TGTCTGTGGC 60
CGCGGCTTC ACGATGCGT CGGAGAGCT GGAATACAC ATTCAGATCC CGATCAGCC 120
CTGTGAGCG CAGAGAGCA GCGCCAGCC AGTTGGGAG GAGCAGAAA CTGCGAGCC 180
TGTGTGATT CTTTGTGGT GGGTGTGCG CAGGAGAG AACCTTGGCA ATGACATGTC 240
CATCTACAC AAAAGGGGCT GATCTGAT CCATACACA CGCCCGTGGC ACATGCTTT 300
CTTCTCGAG TACTGTGTA TCCCTTACT TGTGTGTTG GCGCAGAGC TGTCTGAGCT 360
GCTCTTGTAT TATGAGATG AGAGAGGCC CCGTCTCTTC CATCTCTTCA GCAAGGTGG 420
CGTATGCTG TACCGCTAGC TGTGTAGCT CCGTCAGAC CGTCCCTTCT CGCGCGTGGC 480
TGTGTGGCG ACCATCTTTG ACAGCGCTCC TGGTACAGC AACCTGTAG GGGCTCTGGC 540
GGCGCTTCCA CGCATCTGCG AGCGCGGCG CGCATGCTG CGCTGTGTC TGTGTGTGGC 600
CTTTGCGCTG GTGGTGTGTC TTTTCAAGT CCGTCTGCT CCATCAGAG CGTCTTCTCA 660
CAGCCACTTC TATGAGAGCG TACAGAGCC GGGCTGTGCG TGGCGCGAGC TGTACGTGTA 720
CTCGAGGGCT GACGAGTAG TCGTGGCAG ACACATAGAA CGCATGCTG AGGCAAGCT 780
GGCAGCGCG GTCTGTGGCG GTTCTGTGGA TTCTGTGTA TCTGAGAGC TCGAGCAGCT 840
CGGTGACTAC CCGTACTTACT ACACAGGCT CTGTGTGAGC TTCTATGCGA ACTGCGTGGC 900
CTGTGAGCG CATCTCTCA TGTCACTCT GCTCAGAAA TAAATGCTG ACAGCTGCGC 960
ACAAAAAAA AAAAAAAAAA ACTCGAGGG GGGCGCGGTA CCGAATGCG CCTATTAAGG 1020
T 1021

55 (2) INFORMATION FOR SEQ ID NO: 146;
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1285 base pairs

60

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 146:
GCGACAGGGA GGGCGAGGCG AGCCATGCGG CTTCAGATT CGGTCTCTCG GTGACGGCC 60
AACCCAGAT AGGGATTCG GTTCCTCTCA GTCCGAGCC CTATCAGATT TGGATATGTC 120
CTTCATATTT GATTGGATTT ACAGTGGTATT CACAGATGTC CTACAGTTTT TAGGATATATA 180
TAGGAAACT GGTAAACTG GTTTCTCTCG ATTGGATAT GCGGAAAAA CAACTTCTCT 240
ACACATGCTA AAGATATGCA GACTTGGACA ACATGTCCCA ACATTACATC CCACCTTCCGA 300
AGAACTGACC ATTCTGCGCA TGAGCTTTAC AACTTTTAT CTGGTTCGAC ATTCTCAGC 360
TCGAGAGTG TCGAAAACT ACCTTCTCTC TATCAATGCG ATTGTATTTT TCGTGGATTC 420
TCGAGACAC GAAAGGCTGT TAGAGTCAAA AGAAGAACTT GATTCTACTA TCACAGATGA 480
AACCATGCTT AATGTCTCTA TACTGATGCT TCGGAATGAG ATCCGACGAC CTGAAGCCAT 540
CAGTGAGAG AGTTGCGAG AAGTGTTCG TTTATATGCT CAGACAGAG GAAAGGGGAG 600
TATATCTCTG AAGAACTGA ATGCGCGACC CTTAGAAGTT TTCAATGTGTA GTGTGCTCAA 660
AAGCAAGGT TAGCGAGAG CTTTCCGCTG GATGCGCAG TACATGATTT AACHAACT 720
CACATGCTT CAGATCTCA ACCTTCAGGC TTACTCAGAG ATTGTATGCT TCACATGCA 780
TAACTTGAT TCATAGACT TTGCTGCTTT ATAAACGTA TGTTTTTTAT ATTATATATA 840
TTAAATCAAC TTAAATTTCA TCGAATGCA AACTGATTC AAGTAAATTT GATGATCA 900
ATGTAGCTT TCTAATCCA TAAAGTACTT TGGTTTTTAC AGTTTATAT CTGACATCA 960
CCAGCGCCA TTGTGAAAG GCAACTTTTC AGCAGTACAT TTGAAGCACT TTTTAAACAC 1020
ATGAACTAT AAACCAATTT TAAAGCTCA TCATGTTAAA TTTTATATGT ACTTTTCTG 1080
AACTAGTTT TAAATTTTAT ATTATATGTC CACTATCTT AAGTGTACAG TTAATATATA 1140
GCTTATGCA TGAATGAG ATGCTTACA GTTTTCATA ACTTTTTTTC TTATCGAAC 1200
GTATGCAAT AAACAACT CTATGTTTTC GCAAAAAAA AAAAAAAA ATCGAGGGGG 1260
GCGCGTACC GATTGCGCC TAAAG 1285

(2) INFORMATION FOR SEQ ID NO: 147:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1386 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5 GCGACAGGT GCGCGAGGG TCAGTGGTTC TCTGGGTCT CCGACAGGT GAGCAGCTTG 60
ATGAAGGCA CGTTCCTCAT GCGGACCTG GCGGGTCCA GAGATCTTG GCGGCGCTCC 120
CGAGGGGCG CGAGAGCGG TTACAGTGA TTTCTGATTT TACATGACC TTGAGCAGGT 180
TTGCAATATA TCGAAGGCA TCGCTCTT CTACATAT TCTGATAT ACCAAGTCA 240
TCATGAGA GTGTGGAAG GAGCTGACAG CGTCTCTTCA CCAATATAC CCAATATAGA 300
TCAGGCCA CCGGACCTTC AAGGAGAG TACTCTATAT GTTGTATG TGAGCCAAAG 360
CGCAATCT CTATGTTCAG CAGAGATTC AGAATTTTCA GATAGCCAG GTGTATTAG 420
AGTCAATGC AATGCTGAG GAGGATATA AGACTTCTT CAGACACTC TACCTTACA 480
ACATTCCTCT TTTCACTTT TCTGCGGCA TTGTGATAT CTTGGAAGAA ATTATCCAC 540
AGATGAAT GTTCACCC ACATCTCA TGTGTCTTAA CTACATGAT TTTATGAG 600
ATGTTTCT CGAGGATTT AAGGGCAGC TGATACAC ATACACAG AACAGCTCT 660
TGTGTGAA CTGTGTTAC TTCCAGCAC TTGAGGGCA AACCAATGTC ATCTCTG 720
GAGCTCTAT CCGGACCTC ACATGCGG ATGGGTTC TGTGTGAG AACATTTCT 780
AAATGCTT CTTGAATAC AAGTGTGAG AGCGGCGGA AGCTTACAT GACTCTATG 840
ACATGCTCT GAGAGAGAC GAGACTCTTG ATGTGTCAA CCGGCTACT CAGCAGATC 900
TGTGCGAG GCTTCAGCT GAGATGCA GCGCTGAG GCGCAGGCTN CGAGCGCC 960
TCAGCGCT GTTGAAGAG GCGGCTTCC CAGATCTTC TCCCGCTGA ACACAGACA 1020
GAGCGCGT TGGCAGAG TGGTGGTC CTTCGCGCC CTTCTGCTT CTTTCTCTG 1080
AGCACTTCA TCACAGAG CTTGAAGAA CCGGCGCAT TGGCAGGCA CAGGCACTGT 1140
TCTGTGAA CTTTGACCA CACATCTCA GTGTCTAGG GATTGTAT TCGAGGAT 1200
TTCTGAAA TTTTAAACA TGGAGTTTC AACAATAT AATGTGAA ACAGATCAA 1260
ATTTTAAA TCAAAAAA GCTGCTCTGA TTGAGGGAT GTGGTGGG GTAGAACTG 1320
GACTCTTG CTTGCGGCA CATGGATGC TTTAGGAG ACATTTTGA AACCAACAA 1380
AAAAA 1386

(2) INFORMATION FOR SEQ ID NO: 148:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2098 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

5 AACCCTCTTC CCCGCTCTG GCACTCTGAC ATCTTAAAGC TCACAGCTGC TCCTCTTCTC 60
TGGGTGAAGC CATCTCTCTG ATCGAAGATT CTTGCAACTG ACGAGGAGAG GCGCTCTGAA 120
CCCTAAATGG GTTAAATCTC CTCGACCGA GACTCTTAC CTCGAGATG GCACTTACTC 180
CTGGCAATGG TTCTGACCA TGGGCAAGC ATGAGAGGC ACCCTACTGT GTTATCTGG 240
CCTTATATGA ATGAGAGGC TTGAGGAGA GCGGCTCTGG ATGAGAGAT GCACTGAGC 300
CTTGGCTTAA GCACTGGCT GGAATCTCTT GGTCTATCTG ACTGTGATTC CAGAGAGAGA 360
AACTTGTATT CTCAGGATTC AGATCTTACC CAGCAGCAGA GATGTATTTT CAGAGAGACA 420
GACCTAGAAA TGGGCTCTTC TGGATTTCA GATCTAGCA AGCAGGAGAG GCGCAGGAG 480
CTTCTCTGGG TCTACAGAG AAGGTTCTTG TGAAGGCTAT CAGTGTCTTC CTCTTACTTT 540
GCTGTACTT TTGGGCTCTC GCGCAGGC TGGCAGACT CCTTGCTGT GATGCTATTC 600
TGTGGCATC TGCCTTCTGC CGAGCTCTG CAGATCTGC TCCCTAACCA ACTTCTCTG 660
GCGCTTCTCT GTCCACTGGG CTGATCTAT GTTCAAGCA CTGAGCTGCG AGGCGAGCA 720
CTTCTTCCA CTTCAAGAT AGTCTCTGCG CCGCTTCTCT TGGCATTAAG ACGACTTTTA 780
ACGATGAGCC ATGTGCTCTT TTGGCTTCTT CTCTCTCTTT CCAATCTCTT CCGTCCAGT 840
CACTCTCTGG GCACTATGG ATCACTCTGC CCGCAGCTGT GTGGCAGAC CATCTGTCTT 900
GTCAATCAG AACTGCTCTT GAGCTCAGG CTGACAGCAG ATCAGCCACA GCTGTATGCG 960
TCAAGGCCA CTTTCTCTAC CATTCTCTCT CCGTCTCTCT TCTTTTCCA CAGCAGCTT 1020
ACCTTTTCTC ATCCATCTTC ACATAGGCC CTTTCTCTTT GACCTGAGC CTCCATTTTA 1080
CCTGCTTGA ACTCTCAGC TCTTGAAGA GTGGGCTCCA CTTTCTTCTT GCGCTATGCA 1140
GTCCAGCTTT CACTTCTCAC CGAGCAGGTT TCACTCATCT GCTTCAATGT CTCTGCGGCT 1200
TTCTCTTTC CTTGAAGCT AGCTGGGCTG GTCTTCTCTC CAGCTTGTCT CCGCTCTCTC 1260
GGAATGTCTT TTGAGGGGCC CTGTCTTTTC TCCGAGCA GTTCTCTTGG GTGCGATGCG 1320
AAGCTATCA GGGGCTTTTA CCGTCTTAC CAGCATGTGT AGCAGCTGCG TGCATTTTAT 1380
CTCCATCTGG TCACTCTGAG TCCAAAGCT TCATCCGCA TGTTTTCTGG GCGCATGGAG 1440
GCTGAGCTC CATTCTTGG GATGTGCTT GAGCTGTGTT AACCACTTAT ACCCGAGTG 1500
CTCTTTCTAT GGTGTGCTT GCTATCTTG CTGATGCAA CTAGGAAATT AGGCTGATC 1560
TGGGATGCG TTGCGCTGAA GAGTCTCTTT GCTCTCTCTC AGACTCAGTC ACTGTCTTC 1620
CTCCGCGCTT CTCTTATTC CATGCTGTT TCGAGCTCTC CAGGTAATT TGGGCTTCA 1680

CTCGAATTC CTTTGTGCTT TGTCTCTCTG CTCTCACTC TTGTATTAG AAGCATTTCA 1740
CAAGAGAGA GTGCAATCC TGTCTGCTCT GTCCAGAGCC AGATTTTCTT CAGGGGCGCT 1800
5 GACCTTACC CTCAGCTCA GCGCTGTACA CCGTACCTCG TAAATGAGTG GGGTTTCTCTG 1860
ACTGTATCC CTGACAGCAG TAAAGCAGA AGGACTCTTG GGGCTCAGT GTGAGAGCA 1920
GGTTTACTTA CTCCTCCAG TGAAGACAA CTCCTAGGCT GTATCCATA ATTTCAGAT 1980
10 GAGAACATTT ACAAATAAA ATTGTAGTA ACGTAACTT CATGAGACT AAAAAAATA 2040
AAAACTTGG GGGGGGGCCC GTAAACCAT GGGGCGTTTG GGGGGGAGTT TTAATAAT 2098

(2) INFORMATION FOR SEQ ID NO: 149:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1847 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TCGACCCAG CTTCCCACT GAGCGCGCG CCGAGCGCG TTGAGTCTG GTCTTCTGCT 60
30 CCGGCTGGG GACCAAGCG TGTCTCTGCG GCGGAGAAA GATGTCTCTT ACCGCGCTCG 120
GCGCGGCGAC GCGCGCGCAC GATGAGCGC AGCGCGAGCG CCGGCTTCC CCGAGCAGCT 180
GCGCGCGCTG GCGCGCGGCG GCGCANTGCC CCGCGCGGCG GGGTGAAGCT GATCAGATA 240
35 ATGTTGAGCA TCAACCGCTT GAGAGACTTG AGGTGTACA TCAAGAGTGC GCTTCCCTG 300
GTGCTCTTCA TCAATAGGCT AAGCGCATG GCGATAGCTT TCTGAGCGCT GGGCTACTTC 360
40 TTCAAAATCA AGGAGATTAA ATCCCGAGAA ATGGCAGAG ATTGGATTA TTTTCTGCTA 420
CGGTTCAATG ATTTGAGCTT GTGTGTATCA GAGATGAAA CCGTCAGCA TCTCAGAAC 480
45 GACACAGAA CTCGAGAG TACATGAGC AGCGCGAGCG CCGAGCGTTC CAGCCAGTCC 540
CGCCAGCGCC TGGAGGACTC GGGCGCGGTC ATATCTCTAG TCTCAATCAC CCTAAGCGTG 600
GACCCACTGA AACCTTGG AGGTATTTCC CCGAACCTCA CCAATCTGTA CTCACCATC 660
50 TTAGGCGATC AGATTGAGT TTCAAGCGAG GAGCGCAGG AGGAGATTAA CATTCACTTC 720
ACCTTCTCTA CAGGTGAGG CTCAGATGAC TGGCGCTTCC AGGTCACTG TGAAGAGTG 780
GTATTTACAG CCGTCAATGAC CCGCAGCGCC AGCGCTGGG TGTTCGCTT CACTGTACAG 840
55 CAGCGGAGCT GTTCTCTGTA CAGGTAGAG AGCGCAGCG TCTGTATCAA GATCTGTGAA 900
ACTGCGAGAG ATGCCAGAC AAAAAAGGCG CAGATTTACA ATCCTTTCTG GTTGTATAG 960
60 GGGCGCTTGG GAAGATCTTA TCAATCTTAA ATTCGAGCG TTACAGTGA TTTTCCAGAT 1020

5 GATGACCGTT CATTAATTA TTTCATCTC ATGCACACCA GTTACTTCTT CTTTGTGATG 1080
GTGATACAA TGTTTTGTCTA TGTCTTATC AAGGGCAGAC CTAGCAAAAT GGTCTAGAGC 1140
AATCTCGAAT TTGTGCGGA GAGGTGGCT TTGGCTGAG AGCTTAATCCA CAGCTCCTTG 1200
TTTTTTGAGA GAGACTGAGA GAACCAAT CTTTGGCTGC TGAACTGAGC CTGGGCTGCG 1260
ATGCTCTGCG AATCAATAT CTGGGATCT TGGGTATATC CAGCCAAAGA CATTCTAGAT 1320
GGCTGTAACT GATTGTATCA TATTTATTA AATCTAATCA GAAATGGTCT CAATTAATCA 1380
CGTGTCTTGC CTGGGTATCA CCGAGGCC TTCAACCCCA CTTTGGATCT GAGGACTTAC 1440
CTGATGGGAC GTTTCAGCT GTCTCTAGAG AAGGATCTCT GATCTAGCT GTTCAGGAG 1500
ATGTTTTCAC CAGGTGACA GAGGATGCG GTGCTGATG GGGTTGAGAT TTGTTTGTGT 1560
TCTTTGTTCA GCGCAATATG TAGAGACAT TTGAACAGCT CTGCACCTTT GATACGGTAT 1620
TGCAATTTCA AAGCCACCA TCAATTTCT GCAATTTATG TGTCTGTGCG TTAAATATCA 1680
TAGTAAACAC AATTAATAC TTTTCTCCAT TTTCCTTCCA GGAACATAC CTTAATGTTTT 1740
TTTTTTTGTG TTTTGTGTTT TTTGTGTTTT GTTTTCTTTT ATGAGAGAA AATTAATATG 1800
TACATTTTTA ATACTACCA AATATGACA AAAAAATGCG AGGGGGG 1847

(2) INFORMATION FOR SEQ ID NO: 150:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1569 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

5 GAGCTGTAGC AGAGAGGCC TCTCTCTGA GGGTTGTGC TGTGTGCG TGACCTGGC 60
GGATTAGGCC AACTCGATC CGGCGTCTT GAGTCTGGA CCGATCAGA AAGCGTAGC 120
CAGCGCTGTT CAGGAGGAG TAAATCTCT TTGTGGCTTG GAGCGCTCT AGGTCTCTGC 180
AGAGGAAGCT CTTTCTGGG CTGTGTAGCC CTGTGACATC ATCGACACCA GTGATGAGAT 240
GGATGCGCAG GAGGAGACA TCAATGAGAG AACTGTCTCC AGAAAAAGA AAGCAGAG 300
ACACAAAGA GAACTGAGC GGGCTGAGC AGAAGATAT CCAATGATA TTGTGCTATT 360
GGTGGCTCTC TATATCTCTC CTAGGAGAT TGTGAATTTT TCCCTGATTT GTAGAGATCC 420
CTGGACTGTC ACTTGCAGTG CTGCTTTTG GACCAGGTTG TACCGAGCA CTACAGCGTG 480
GATGCTGCC TGGCTTTGCG TTGTGACCA GATCAATGCG AAGAGTGGCG CTGTCTCCGG 540

5 GCTTGTGTCA TCCGATCTCT GTACATATG TATGAGCCAT TTGCTGCTCG AATCTCCAG 600
AATCAGCCA TTCCAGAGAG CAGCCCGAG ACATTAAGA ATTCCAAATG CTTACTTTTC 660
TGTGTGAGAA AGATTTTGG GAGGAGAG GAGCAATGT GCGAATTTCA CTTCAAGTTC 720
AAAAAGAT CCGTAGGTT AAGAGCAG TGTACAGGAG GATTCAGGCG TCCGTTTCAG 780
TAGAGAGATG TTCAATGCAA TCGAGCAG GAGTCTGCG TACTCAGGT CACACGCTTC 840
AATTTCACT TATTTCCGAT TGTCAATGGA ATGATATTTA CTGCTTTTAC TATCAATGCG 900
AGCAGGACA TCCGCGATCA TCGAGTGAGA CTGCTTTTCC AAGATTTCCC TGTCCATGAT 960
GCTTCGAAAC TCCGAGTGA ACGGCTGTG CAGTCAATCC TGGACCCGAT CACAGCGGTT 1020
CGGCTCTTTG ACTGTTGACA TCTCTAGTAC CCAATCTGCC TGAAGCGTGA GTTACTGCTT 1080
CCCATCGCTT GGGGCGAGCC TCGAGTGTAG TCCATTAATA ATCAATATCC AATTTGAGCA 1140
GGTGTGCTGG AATGTATATC TCGTTAATGA TGTACATGCT CTTCAAGGTTC TAGGGCTCTT 1200
GTTAGGGGAG GAGCAATGT TGAATCAGA GCGAAGACA CTACTATGAT TTATTAACAT 1260
ATTTTAATGT AAAAATTTGC ATTTAAAGG AATGCGGCTG TTTTCTGTGT TAAACGCCA 1320
TTTGTGCTGA TTGATTTGT TCTTTATCT TTATTCGCG TGAATATGT TCAATCTGCT 1380
GTAGGAGAAA ATTAAGCTCT TTGATCTCC AAACAGGAA GTTTCAAGAT TCCCTTATGG 1440
ATCAGAGGAA CTTTAGAGGC CTGAATTTGT TCGTTCCAGT TTAGCTGCCC CTCGAATTCA 1500
AGTAAATAT TTCCCTTCTC CTTTAAGCT TCTTCAGAA TAAAGCAAGT GACAGGTTTT 1560
CAGAACTTT 1569

(2) INFORMATION FOR SEQ ID NO: 151:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

5 CCGACCGTCC GCGAAGGAT GACAGATTAA CCAACATCTT AGCCCCCATG GCTGTGGCC 60
AGATTATGAC ATTTGCTGCC CAGTCAATCG GCTGTGGCTT TATTTGCGGA TGGAACTTGG 120
TATCAATGCG CTTGAGTAC GTCTGCTCT CCAAGGTTTA CCAAGAAACC CCGCTGCTAG 180
CTGTGAAGCC TGTCTTTAAA GAGAGGAAA CTGAATTTGA AAGCTGAT TTACACAAAG 240
ATACTGAGCC AAAAGCCCTG GAGGAACTC ATCTAATGCG TGTGAAGAC TGTAACTGCC 300
ATGAGCTTGA AATGAGCAA GAGCTACTT GTCCCTCCA GATGGCTGAG CCTTTGCGTA 360

420 CCGTCGAGA TGGATGGGTC TCGTACTACA ACCAGCCCTGT GTTCTGGCT GGCATGGGTC
480 TTGCTTTCTT TTATATGACT GTGCTGGCT TTGACTGCAAT CACACAGGG TACGCCACAA
540 CTCAGGACT GAGTGGGTC CATCTGCAAT ATTGTGATG GAGCATGAG TATTAATGGA
600 AATATGGAA CTGTACTTTT TACTTGCGTA CCGCGAAT GTGCTTTGGT TCGGAGGTC
660 TGAATGAGG ATTGGCAGAG GTTCTGTATT TGAATCTGTS TGTGATCTCT GTATCATGC
720 CTGGAGGCC CCGGACTG GTGCTTTCTC CTTTGGAGA TATCGGATCA AGGTTGATTC
780 AAGGAGGTC AATTACCACT ACCAGATGAC CTGAATATAC AACTGAATTA TACATCTCTA
840 ATGGGCTTAA TTCTGCTTAA ATTGTCCGG AGACAGTCC TGAATCTGTG CCGATATCT
900 CTGTGATCT GCTGTATTGA GGGGCAATG CTCTAGAAAT GGGTCTTTGG TCTTTTGAT
960 TAACTGTGAC ACAGTTGCTG CAGGAATG TAAITGATC TGAAGAGGC ATTATTAATG
1020 GTGTACAGAA CTCATGAA C TATCTGTG ATCTTCTGCA TTTCATCTG GTCATCTGG
1080 CTCGAATCC TGAAGCTTTT GGGTGTGCTG TATGTATTC AGTCTCTTTT GTGCAATGG
1140 GGCATATAT GTATTTTGA TTGGCCAAA ATACTCTGG AAACAGCTC TTGCTTTGG
1200 GTCTGATGC AAAGAGAT AGGAGGAA ATCAGCAA TACATCTGTT GTTTGAGCA
1260 GTTTACTGT TCGTATCTG TTACTAGAT ATATAGACA CATGTCTTA TTGTGACTG
1320 CAGATCTCA ATAAATGCT GGGTGTGTT GTCTGTGTTT ACCAGCTG TCGCTTGA
1380 ACTAAGGCT GTTTAGAAA CCGATGAG CAGAAATTA CTGATTAAT TTCCCTATG
1440 TTGAGGCTCA TGGRAAAA ATTGGGAAA GGAATTAAT AGGAGACT
1500 ATATGGAAT ACATGATTTT CCGATTTTC TCAATGATG ATCAATCTT ACGTAAAGA
1540 GTGGTATGC ACCTGAATC AGTTATCAT TTGACAGATC

(2) INFORMATION FOR SEQ ID NO: 152:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1719 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

60 TACTATGAG GTCAATGGA AATAGACA CAAATTTACT GGGCTAGCA TTTCATTAAT
120 TACATGGC ATGATAGGC TTGGTTCAG AACTTGGAT GATGTGGCT GCTGGGCGAG
180 TAGAGCAAT GTCTAGATC ACCTTTCTG CTGTGATGCT ACTTGTCTCA CGAATCTGT

240 ATCTGATCA ACAGGCTGTC GTTCAGGAA TGAATACAGG AATTGAGGA TTATCGAATG
300 GTCTGGAGC GGGCTCTTAT GATTCATATT TCTATATATT CAAATGGA CTTAAGAAC
360 TGGCAATAC AGGACAGAC TTGGAGCAA ACAGAGCCC TACAGCAC TTTTGACAGA
420 ATTGCAATAT CCGTGGGCTT CCGTCTCTAT TTGAGGCTG TTCACTACTG CTGGCTCTGC
480 TTCTGGCTT GTTATTTCC GACATACCA ATTATAGCTT AAGTTCGAG AGTTGGAGAA
540 AGCACTGTG CAGTCAAGC CATCTGATA ATACAGAGC CCGAGAGAG GCCAAGAAC
600 CTTTACTCCA GGCACAAAT GTGTGAGAC TGAATCAGG AAGATTTTTC TATCAGAGCC
660 CAGCTCTTAG TTTCAGCTC TAGTCTGGA TGTATTTCC AATTCCATCC ACAGTATAT
720 TTATATCTT CTTAAGAAAT GTATCTGCAT GACTCCGTG GCACTTAAG GAGTGGGAA
780 CTTAGACCA GACAGTTTC CAAGATGTT ACATTTCTT TTGAAAACC TTITGTTTAT
840 TACCAACAA TTCTGCGAC TACGTATTT GTTTATATAT ACATCTCTTA ATTAAAACT
900 ATATATGAA CTCTTAGAT ATTAGCAAT GTCTCTGCTA CCAATTTCTT AAGGTGTGCA
960 GCTTTACTG TATCTGACT CAGTGAACA CAGTAGTAG TATGTTGTG GACCATTTG
1020 TTATAGATT GTAAATTTT GAGTCAAT TTATATGTT AAATCTTGG GTCAATTAAT
1080 TCAAGCTTT AATCGAGTG CACTTAACA AAGAAATGT AATGAAATG TTTCATTTA
1140 AAAAAAAA CTCTTAGAA AACTGACTA AATCTGAAT ATGTTTTGAG CTGTGTTGA
1200 GTACTTTTAA ACATATCTA CTACTGTTT TGAATGAGA AAGTATCAG CATTTAGAT
1260 TTAGTTGGG GTATTTAG CCGTAAATCT AATCTGGC TCAAAATTAAT TCCCAAGCTA
1320 CTCTTATAC CACTATCTT TTATGTTTG CAAATGATA AGCACTGCA CACTTGATA
1380 CATAATCTAA AATATATA GTAAAGCTG TACCTTGA AATCTGAGT TGAATCTAT
1440 TATGTAGATA AATATATA GTGGCTTTC AGCACTGCA CAGTAACTCT TTATTTACAG
1500 ACCTATGTT TGTCTTAAT TTTCAGGAC CTAGAGAGA GCTTTATACA ATTACCAATG
1560 TGAATTTCT TAAAGTGTAT ATTTTGTGT CAGTTATAT TATTTAATA AGTGTACTT
1620 TGTAAATAT GTATTAAG AACTATAG TTTCATCTG TTTCATCTG TGTGTGTTA
1680 TTCTTATG CTTTATTAAC TTGGAACT CACTATGTT AATTAAGCTC TTAAGAGAA
1719 TGTAAATAT TTGTATTA AGTTAAATAT TTATATAT

(2) INFORMATION FOR SEQ ID NO: 153:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 863 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

5 GGCACGAGG AGGCGGGGAC GATGTCGCA TCACACCGA GTTGAGATT TCGAGTCT 60
TCCCTTAGAG CAGGGGAAAC AGCTCTCAT CAAAGGAAC AGAGGCTCT CCGTCAGTGG 120
10 TAGCGAGCA GCGAGAGCG GTTCTCTGG ACTGTGGA TGTGCGCTTG GGGGCGCGAG 180
AAACACGAAG CAGATGCTTC CAGACAGTA ACTACAGCT GGTGCTCTCT CTGAGTTCC 240
15 TCGTGTGTC CTATGATCT TTGTGCATT CTCTTCGA ACTGTGCA MAGACTCTG 300
TCATCCAGCT TGTGCTCTC ATCATCCAG ATATTCACT CCTCTTCAC ATCATCACTA 360
TTTCTCTCAT GTTCTTCAC ACTTCTCTCT TCGAGGCTGG CCTGTGAC CTCTATTTC 420
20 ATAGATTCA AGCGACATC ATCTGACAG CTGTACTT TCCCTGACG ATCTGCTTC 480
ATGTCTGGT CATGAATTA CCGTCGAAA ACTCGACAG CTTCATATG KCMAGTGGAC 540
25 TTCAATGCT GTTGTATC CAGACAGT CAGCATGTT GTACTCTAC TTCTATTAC 600
GGACAGCTT AGACTAGG GATCTCACT TCTACAGGA CTCTTTGTG CTGCGAGG 660
AGTTCACTCA AGTTCGAGG TGACTCTTG TCACATGAT GATACTTTT CCTTCCTGA 720
30 TAGAGGCA CATTGTGTC TTTCAGGG AGATTGGG CATTGCTAG GCGGAAACA 780
GGTGGATTT TCCAGGAA GGGTTGAAA TTAGGCTGT TGTTCAGCC ATTTCAGG 840
35 AAGGGAGG GTTTCCTTC CCT 863

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

40 AACGAAAA AGATGATTT TCTCTGAA TTGTGACA TGAGATCA AGTTTTATT 60
50 TTGTACTAG CTCTGAGG AGATTCAG TTCAAAAG CCGATCTT TCTCTGAG 120
CGAGGCTG CGGTGATC AGTGCATC GAACATCAG ACGANTAG AGGAATTA 180
55 GATGATAG ATGAGCCCA AATTAAGTT TTCTTGGGG AGATGCTCT GATTAAGA 240
AGGTGATC AGCAGGAG AGCAGTGG CAGCCACAC AGATGATCT GTGTGATT 300
60 ACGAGAGCC AGCATTTAG CTTAAGTCC TTCTACTAC CTGTATTGG CATTTAGGT 360

CGAAACCT CTACTGCC ATAGCCAG AAGGTGAA AGAGACCA GTTCTTTAA 420
GACTGCCAG CAGGCTTGA GCGTTATCT ATTAGCTGA GTGAGCAG TACATATCT 480
5 TGTCTCTTT CAAAAGACT TTCTCTCT AGTGAAGTA CTCTTCTT AGTTCAGGA 540
ACAGCTGGA CAGGCTCTG CTGAGTACT CTGTATGAC AAGGCTTGG TTAACTGAG 600
10 GTATCTCTA GGTGTGAGG TTATTATGTC CCGAGGGA ACACATAT TATATATTA 660
ATCCACTTT ATATGATAC ATTAAAGA AAAAAACA AAGGCTTGA AGTGTGAGG 720
CAGGCTCT GATATTGCA GCTGATTTT CCGAAGGA ATCCAGCA ACTGCTTCC 780
15 TGTATTGTT TCTTGAGG GGTGATCTA GAGCTGAT CTATCTAGA TGAAGAGAG 840
CAGCCAGG CTGTCTGGA TCAGCAGCA GATTTTAAA GAAAAAGA AGGTTTCTT 900
AGTGAATTA TTGTATTGA AGATGTCAG TATTAACAC TGACAGATG CTATCATAC 960
20 ACTATGTC GTTTTATGA TAAGATTAC ATATCATAT TCTTTTGGG AATATTGTTA 1020
TTGATATA AAGACAGAG ATTATATTA AATATAAA GAGCCTTAA AAAAAAAG 1080
25 CTCTGCCA ATCTGCTCA G 1101

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2031 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

40 GATTAAGCC GTTTCAGGC TAGGTTGTT GCAAGCCC NCGCTAAG TTTTAATG 60
CGAGGCAA GGGCTGAAA GAGGGGAA GCGAGGGA GCGGAGGT AGCAGTGA 120
45 TTCTAGGCG TGAAGGTT AGCAGGCC TGTGCAATG CCGTGTATC AAGACAAAC 180
CAGGTCCTC CTGGTGGCT ACCAGCTTG GTTTTACA ACGAGGCTG GAGTCTATT 240
TTTGTAGTG AGATACATA CACTTACTG TGGGCCAGTA TTGTAGCTG ACTCTGATT 300
50 GTTTACTAG ATCGCTTCC TGCCACAC AATTTGTGA CTATCTCTC AGATGATAC 360
CAGGCTTTC CCGAGTCCC AACAGAGC TGTGTCAGG CCGTGTATT ATGTCAATT 420
55 TAGGTTTT ATATATGAC TTGATTTCT GTTTTGTGA TTTTAGACA GTTATGAC 480
CTTATTTA ATACATCTCT GTCATACAG ATAGCATAT ATGTGTGTC GTATCATAT 540
ATCTGTATC TGTATTTTC AGATTTGAG CTGAACAGA TGAATCTCT CAGCCAGGA 600
60 GACAGCTCC ATCCCTCTA ATATGTTTG CCAAGATAT TATGATCTT TCTTATTAA 660

5 TATTTTCAT TCAGAGACT GAGCAGAGC TGATGATTTT TGCTGTGTTCT TTGGAGACTA 720
AGTGAAGGTC TTGGATGAC TTGCTGTGTT CCTGAAGTC GACTTTGGGG CACTCTCTCC 780
AGTATTAAGC GCGCTTTTTC CTGTTGGTGA CTCTCTCTCT GCTCTCTCTCT GTGTGTGATA 840
GTGACTCTTC CATGAGTTC ATGTCTGTTT TTGTGGCATTT GGGGATTAAT GCTGAAGCAG 900
10 AGCATTTTGA GTTTTGTGGA GCGCTCTTTC CCAATGATAG ATGACTCTTC TTGACTTGTT 960
ATGTCTCTTT GCTGTCTCTT TTGCTCTCTT TTCTCTTGGA AGAGGAAAGG ACTCTGTGTA 1020
GGCGCAGGCT GAGTGAATG AGCTGAGCT GCTCTATGCG GTTCTTGAAG CAGAGAGAGG 1080
15 AGTATGTGAT TTACTATGAT TCTTAACGAA ACATTTATTC AGGCAAGCT CTTTGCAGAT 1140
CCAGAACTTC AGGCAAGTAA GGGTATGAC TTGCTGATGA ATATGATGCT GCTAGAGGCT 1200
20 AAATGAGGTC ATGAGATTT CTGTTCAGCG GGCAGGAATA GCTCTGAT TTCTAGACT 1260
TTTTTTTAA GCAATTAATT TTGACTTGT TCTCTGTGAA GTGAGAGGG GCTAGACTTT 1320
TCCCAATGTT AGACTAGAT CTGAGATG CAGCAGCTG TATGATTTCT GTTTTCCGAA 1380
25 GAGGAGAAC TTTTGAAGC CAAATGATCT TATTTGTTAT TGGCAGCCG TGCTTTTTC 1440
GGTTAGAAA TTGACATGAG GATGATGTT TTAGAGAGAG TCTTTGGAG CATGGGTTAA 1500
30 CAGGAAAGGC TACTTAATCT CACTATCTG CAGCAGAGC AGCAGCAGG CATTTACTAG 1560
CAGCAGGAAA ATGATGTAT TTGATGTTCT GTGTGTGCA AACTGAGGCA CCAATTTCTT 1620
TCAAAATG CCACTCTAG GCTGGGGCGG GTGGCTGACA CTTGTTATTC CCAGCACTTT 1680
35 GCGAGGCGCA GCGGGGGGGA TGACCGGAGT GCGGAGTTT GAGACAGCC TGAGACAA 1740
TGGAGAGAAC CCAATCTCTA CTTAAATTA CAAATTTAG CCGGCTTGT GCAATGCGCC 1800
40 TATATCTCA GCTACTTGG AGGTTGAGG CAGGGAATTT GCTTGAACCC GGAAGGCGG 1860
AGGTTTTCGG TTGAGTTGAG GATGTGTCCA TTGCACTTCC GGGCTTTGG GCAACAGAG 1920
45 CAAATATCC GTTCTCAATV HTGTGCAAT TGGAATGCA GCTTATGCA ACCGTGACC 1980
TGGAGGCGGG GCGCGGTACC CAATTCGCC TATAGATCT GATTTAGAT C 2031

(2) INFORMATION FOR SEQ ID NO: 156:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1981 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(41) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

5 CTTGCACTCT GAGGCTTTCA GCGCTTCGAG TTGCGCCGAG GTTGGCTTCC TTGCAATCTT 60
TTTCTGGTGA TCAGGTTTTT ATTGAGAGAA CAGCGCCCTC TTGTGACAG TCATATATGA 120
20 GCTCACTCTG GAGAGGAGC TGGGGCAGAG CTGCGCTTGA TGTTCAATTT TACCGCCGAA 180
GAGCGGAAA ACTCTGTCTG GACAGAGAT GGAAGCTCAG CCGAGATCTT GAGAGAGCTT 240
10 GTGTGCTCTG CAGCTGAAGC CTTGCCCATG TTAGAGAGC AGCTCATGGA TCCCGCGGGA 300
CCTGGGGA CA TGAGAGAT GTTTCGCGCG CCGTTGAGCA TTTAGAGAT GCTGATTTCC 360
CTGTTCCTTC GCAATATCC GCGGCAAGCC AGGCTTTCTG ACTGCGAGT TGCGCTCTTC 420
15 TCGCGGGGCC TGCTAGCCA GCGGGGGCCC TGATCCCTGA TCCGCTCTCT GGTATATGAT 480
CTTCTCAGC TCTATCTGAC GAGCTCAGG GAGGCGTTTT GGAATCTGAG CTTTTCCTTC 540
TATGAGCAGC ATGATGAGGA GGTGATTTGT GTCTCTGGA AGCGCAGCAG CTTTCAAGCG 600
20 CAGCGCTTCA AGGCTTCAG CACAAAGGCG GCAATGCTGA TGTCTCGAGG TGGGGAGCTA 660
GTATGCTCC CCAATGTTGA AGCAATCTTC GAGGACTTTG CTGTCTGAG TGAGGCGCTC 720
25 GTGCAAGCTG TGAAGGCGCG AATGAGAGG TGCACTGTGT GATTCAGCT CTGAGCGAG 780
CTGTAGAGCG ACAGCAGGAC ATTGAGCTTC TAGAGAGGA TGTGATGAG ATGAGCTTCA 840
30 CCGCTCTTGG AATGATATCC TCAATGAGG GCGTCTGAG TGACATGCT GAATCATCTC 900
CAACAAAGCC CAGCGCCGAC TTCTCTCTCG ATCTCTGAGC ATTGGGAGG GGGCATGCTG 960
GCGCATGAG TCTCTCTGAG CTAACATTC CAGAGAGGA GTTGGAGCCA GCTCAAGAAA 1020
35 GGAATGAGC CCAAGAGATC CATCAACTTA TTAGCGCTCG GCTTGAAGCT CCGTGGATTT 1080
TCCCACTCTT TTCTTATGCT TCTTTCAGAA ACAGAGAGG GATGTGTCTC CTGGAGAGG 1140
40 CTCTGTCTCC TTGCTGTCTC CAGGAGCTGT GCTTAGACTT AGCATGCGCT TCACTGAGT 1200
GTGAGGCTTT TAGATGAGAC CAGCGGAAA TTGTGGCTTT CTGATGACA TCAAGCAGAC 1260
TGAAGAGTGG AAGGGGCTTA TATGTGTATG ATTAGAGCAC ATTGAGAGG CACATGCCC 1320
45 TCTGTGTGTT ATGCGCACTC CAGAGGTGGA GACATGGA AAGAACCGAG GACAGAGAG 1380
GATTTGGTAT GTGAAGGGGT CAGGGGAGTG GTAGTCAGCC AATCTTGAG AGGTGCAAAA 1440
1500 AGCACTGGGG GCTTACCGTTT AGCTGATCTT GCGTGTGCTG TTGCGCGCTT CATGTGCAA 1560
50 ACTGCGACTA CTATGTACTT GCAATGCGGT TGCAGAGATG GGGGAGCTC AAGTCTTACT 1620
CCCGAGAGC TCCCAAGGCG CAGAGAGGAG AATGCTGCTT CTTTTCAGTC TGCTCTACAC 1680
55 CCACTTTCTG GTAGCTCTTC TGCTCTCTCT AATGTGCTT GTTTTTCAG ACTGAGCTCA 1740
AATAGTCCC CTGCTTACG CCAATCCCTG CCGCGAGGCT GAGGTGATCT TTGCTCTCTC 1800
TGAATATTA GAGCAATTAC TGTCTGTTCA GTTGTGTTG CAGGAGACA CAGTGGGATA 1860

AAATCTATTG TTTTGAACTC TGATTAAATA TTAAATTGCA GCTGGGGGTC GTGGCTCATG 1860
CTTGTAAATCC CAACTATTAG GAGTNAGAG GAATCACTTG ASCYCAGAG TCTGAGACCA 1920
ATCTGGGCAA NAGAGAGACC CCATCTCTTT TAATAAATA GTTAATATTGC TTAAATAAAA 1980
A 1981

(2) INFORMATION FOR SEQ ID NO: 157:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 915 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

GAATTCGCA GAGCGCGGC CATGGCGCTC CTGGTTTCGG TCGTCGGTGT ACTGCTGGGC 60
GGCTCTTCG GCGTCGTGGG GTTGGCGAAG CTCCTCGAGG AGATCTCGGC TCCAGTTTCG 120
GAGCGGATGA ATGCGCTGTT CTGCGAGTTT GCTGAGTGT TCCCGCTGAA GGTATTTTGC 180
TACGAGCGAG ATCCGCTGAA CTACAAATA GCTGTGGCT TTCTGGAACT GCTGCGTGG 240
TTCTGCTGCG TCAATGGGCC ACCGATGCTG CAGAGATCA GTAACTTGTT CTTGATTCG 300
CTCATGATGG GGGCTATCTT CAGCTGSCA GCTCTGAAG AGTCACTAAG CAGCTGTATC 360
CCAGCCATGG TCTGCTGCGG GTTCTGCTG CTGCTGATG TCGGCGAGCT CTTAGCCGAG 420
ACTAAGAGCG TGTCTAGACC CACTAGAGAG AAGACTCTAA GTACATTCOA GGAATCTCG 480
AAGTAGAGCA TCTGTGCTCT TTTATGGCAT GCACTGTGCA CAGCGAGAC ATGCTAGAAC 540
ACAGAGTCTA TCATCTGTTT ACCAGTAAAT TATCCAGGCT CAGCCAGTGT TGAAGAGAG 600
ATTTGTCTTA CTGGGCACTG CTTTCTCTTT TTAGCTTTAC TACTCTTTTG TGAAGGTAC 660
ATGTTATGCA TATTAACTTT GCTCATGCA TATCAAAATA CAATAATAGC AGAAAGAAA 720
TTTAAATCAA CCAAAATTTCT GATGCCCAA ATACCCACTT TTAAATGCTT GGTGTAACTA 780
TACTCTGAA CTTTATTCTG TCGCTTTTAA CAGATATATA TTTTITTTTA ATGAAATATA 840
AACCATATAT CCAATTTTAT TTCTGCTTTT TAAACCTTA TAAACTATTA TAAATAAAAA 900
AAAAAAAAA CTCGA 915

(2) INFORMATION FOR SEQ ID NO: 158:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2117 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

AGAGCGAAG GAGGTGGCG CTGGTCCGCG CATGAACTG GCGCGGCGCG TCGTGGCGCT 60
GCTGCTGCTG GCGCGCTCGG TGGTGCAGGC GGTGAGGCGC ATCAGGCTCG GACTGGCGCT 120
GGCGCGGCTC CTCAGCGCTT ACHTCTACCC GCTCTCTTAC TCGTCTTTG CCGATGCTG 180
CGGCGAGAG CGGAGCTTTA GCGCGGAGGC ACTCGAAG GATCTGAGC ACAACTCTT 240
TGGACGCAAT CTTGCAAGA AATCAATCTT AATGCGCTG TTGTTTCTA TAAACAGCC 300
AAGGCGCAG AAGCTCTCA GCTCTGCGCT GCAGCGGTG ACAGGAGCG CCAAAATTT 360
GCTGCGAAG ATCATGCGAG AGATATTTTA CGAGGCTGT CTGACAGTG ACTATGTCA 420
CCTGTTTGTG GCGCAATTC ACTTTCACA TCGTTCMAAC ATCACTTGT ACAAGATCA 480
GTTACAGTTG TGGATTGAG GCAACGTGAG TCGCTGTGCG AGTTCATCT TCAATTTGA 540
TGAATGTGAT AAGATGCAAG CAGGCTCAT AGATGCCATC AAGCTTTTCC TCGACTATTA 600
TGAAGTGTG GATGCGTCT CTTACAGAA AGCAATGTC ATATTTCTCA GCAATGCTG 660
ACAGAAAGG ATCAGAGTG TCGCTTTTGA TTCTGAGAG AATGCAAGC AGAGGAGCA 720
CATCAGCTC AAGCAATG ACAGCGCTT GTCTGTGTG GTTTTCATA ACAAGAGAG 780
TGGCTTCTG CAGAGCACT TAATTGAGC GAGCTCATT GATTATTTG TTGCTCTCT 840
CGCGCTGGA TACAAACAC TAAAAATGTG TATCGAGTG GAATGCAAT CCGAGCTTA 900
TGAATTTGAT GAAGCAATG TAGCAGAGT GGTGAGAG AGTCAATTTT TCCCAAGCA 960
GGAGAGATT TTCTCAGATA AAGGCTGCA AAGGTGTTTC ACCAGTTAG ATTATTAATA 1020
CGATGATTA CAGTATGAT TGGAGCGCG AGTCACTGCC TGGAGTTGGA AAGAAACAA 1080
CACTAGTCC TTCCAGACTT CGAGCGCGAG CTCCTTTTCC TGGAGAGCA ATCCAGTGA 1140
TGTGCTGTT TGAATGTACA GGAATTTTCC CTGGCAATGT TTCCAGCGCC TGTGCTGTC 1200
AGCGCACCA GCGACGAGG GCGAGAGCT GAGGCTGCC GAGCAGCAC AGAGAGAGG 1260
AGCGAGTCC CAGCGCACTC ATGCGAGGC TCATGATTTT TTACAAATTA TGTTTTAAT 1320
CGAGTGTTT CTGTTTCAG GAGGATGAA TAGTTTATAT TGAATATG GATACATTTAT 1380
TTAAATGAT TTTTAACAT ATGAGAGCT GCTCAGATTC TAGTTGTG GCTTTGTG 1440
TGTGTTTTTT TTAAATTTCT CATCATTAAT ACATGAGTG TATGTATCT TTACTGAAA 1500
TAGCGCAG CAGCATGCA TGGCATTTGT TCCAGAGAG GCAATGCGT GGAATGTGCG 1560
TGGACATCA GCGAGTGTG TCCAGAGTG GTCCAGCGG ATGCTGCCAG GGCATTTGA 1620

5 GTGTTTAGGT GAGGACAG TAGGTAGAG GAGCGCTTCA GGCACACAG ATTAGCGTGA 1680
AACAGCTCTT CGAAGGGTTT TCACTTAGC AGCAATGGA GCTGTGGAG TGAATTTGCG 1740
CAGACTGTCA ACATTTGTGA GAGCGTCTT TTTGAAAGA AATATTTTC AACTTTGAC 1800
TTGCGATCA CTCGTTTTC CAAAGTGGG CGCTTCACT TCCATTTCA ATAGCCACA 1860
CGTGCCTCT GCTGATCTT AATATGTC ATTTTGCCA TATTAATCT TCCATTTTA 1920
TACTATAT TGTAGTCTT ATTCAGATC CCGAAACT CCTATTAAGC TTAGCTGCCC 1980
CTTCTAGCA TGTCTAGAC GGTGTCTTC TTTATTAATG CAAATGCTA CGTTTTACA 2040
15 ATAAATTTT GCAATGCA AAAAAAAAA AAAAAAAAA AAAAAAAAA GGGGGGGCG 2100
GTAACTATT TGNCCG 2117

20

(2) INFORMATION FOR SEQ ID NO: 159:

- 25 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2195 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

35 TGTTCCTTAA TCCCTTTTCT AAAAAAGGG GAAATCCCG ATGCAATTTA GCGATGTC 60
TGTGTGAG TGTGTTTAT TGCACCTA ATCTGTAT ATAGCTTTT CATTCTCG 120
CAAGGCTTT ATTTTGGAG TTAGCCAA TGTGTTTTC AAAAAATTG TTAATTTCTC 180
CTCTTCTT CTCTTCTTCT CTTCTTTT CCGCTCTGAC CCAAAAGTT ATTCCTCAA 240
40 CAGTACTGA CAGAGCTTT TGTTTCTGA CCTGTAAAT TCAAGTCTG CTAAATTTGA 300
CAGAGTGC AGTTTGTGG TTAATGTCT GATTTGCTT ATCAATCAT ATTAGCAGA 360
AAAAAAGA CTGTGTTCTG TTTTACTGA GTCTTAAGA AAGTGGCCC ATAGTTTAT 420
45 GCAATTTT CAAAGCTTT AGTACACT GTATTTCAA ATGGGGGAC CAACTCCCG 480
GAGAAACA GCTCTGAAC GACTAGTGC TCACTTAGA AAAAAAAG ATCTCTAGC 540
50 TGACCAATT GACTTCAGA TGTATGTC CTTGTATTC AAGGAGAGA AGAAATGTC 600
AGCACTTTT GAAGTCTG AGTATPAC AGTCATGCA AATATATG AAGAAATAT 660
CCTGAAGGT GTCCGATTT CAGCTATC CTGGAAAT TCCCTAGAGC TTTTACAGA 720
55 GCAATGTA CAGCTCATG CTCTCGATA TCACTTATG AAGAGGAT TAAITGCTG 780
TACTGAGAG ATGCAATTA TCTTTGCG TCGAATAT ATGAAAAA TCCCTCTCT 840
60 CCTGTTTCT AGGTGGAAG ATCTGATCA GCTTTTAGG CCTTTGAG CAAATTTGAG 900

5 TTTTCACTG GTGACTATG AAAAAATT CTGACTATC TGACCCGCA GCAAGAGCT 960
GCAATGTTG TCACTATGC TACCACTCA GTTTTCTCT TCAITGACG ACAGACTTG 1020
CAGACTCAA AAAAAAGC TAAATCTTC AATATGCA GCATCTGCT CTACTGCCA 1080
CAGAGAGC TCACTGCTG GGGGATTTG CACCTAGAG GTCAGCTGC GTCTTATAT 1140
10 GGCAGATAG ATGACTGAC AGCAATGTC AGAATGTCG AGAATGTCG AGCAATTTT 1200
TTTCTGTTT TCTTACACT TATCTCTTC AGATTTTAA GAAATGTCG TCACTGACG 1260
AAGCTATGC ATTTTGAAC TTTTCACTC TGGATTTT TAAATCATTT TTAITCTGA 1320
15 ACTTAACAA AATATGATG TGTGCACTG ACTGTGCA AGAGATGCT TTGCAATTT 1380
CCTGCACTG ATGATATCT TACTAAATG GTGAATGAA AGCACTATG TACACTGAA 1440
TGCCTAATG TATCTGAAG CAGAGTGA TACTCATTT TATGCTCTC CCAITGTCG 1500
20 TGTGTTTTC CTCTTGA CA TGTGATCA GTATTTAGAG GTGAGAGT GATGTATCA 1560
GATTAATTA ACATTTTAA AAGCAATAC TTTGCTTAA TCAAGTGTG TCAAGAGAC 1620
25 TGTGATAC ATTTATATG CCGAAGTGG TTTAAAAAA GAAATATCA CCACTGGAA 1680
GAAATCTTAA ATGAAAGCG CATCTCATG TAGGCAATTT TCCCTCATAT TTACTGCGC 1740
30 CATGTTGTT TCCGTGACT CATGTATTT TTTTTCGAG ATCTCTTTC CCAATGCT 1800
ATTGAGAG TATCTGCTG CCGTGTGATG CATTATACA CATTAAAGCA GATCTGAGT 1860
CTGAATGAG TATTAAGAG CTATTAACA GAAATATG CATTAGTGC AAGAACATA 1920
35 TAGTGAAG ACTTTCTTT TGTGTTGTT TGTGTTGTT TGTGTTGTT TTTGTTTAA 1980
CAGAGAGAG ATTTTATTA CAAAGAAAA AATTCAGT AATTGTGAG AATGCTGCT 2040
40 TTTTACCA TCTTAAGA AACTTTACA AGGTGTTTT GAGTGAAGA AAGCTTATA 2100
AAGTGGAT CTAAATGTT AAAATTAAC ATTCAGTGC AAGTTCTTA AAGCAGACT 2160
45 CATTTGTC ATGAGATA AGAAGAGCT ACTGTAGAG TTTTTTTTT TTTCTCTTT 2220
AATGAGAA AAGCTTCTG TAAAGTTTC ATACTTTAT TCGATTAAT CTGAATGATC 2280
CTACTCTTT GCAATTAAC TATGCTTAC CAGTTTCCA TGTATTTAG CTTCGTTG 2340
50 GATTTGAA AAAAAAGA AAAAAAGA CAAAGCTTA ATAAATAGG TGAA 2395

55 (2) INFORMATION FOR SEQ ID NO: 160:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2120 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

5 CCGCGATAC CGCTGAGCT AGTCCGATC ACACCTCTCG CGTCTCGCG CGTCCGAGCC 60
TAATGAGAC GCTTGGGAA AGGAGATAC GGAATTTCCG GTGAGCTTC GCTTTAGCGC 120
TGTGAGTTC TTCCGCCAA CCCAGGAA GCGGAGAGC AGTTTACAC AGCGCCGCTC 180
GTGTTTACGG CGGCGCCCGC TCGCCGCCCA TGTTTCTCTT TTCTCTGTT TTCTAGAGT 240
GCTGCTGTA ACCCGCTCC CGGCGCGAC CATCTGTTGC CATTCGCGCC GCGCGAGGCA 300
TTCCAGATT TTGAGATCG CAAGTTTCAT GACACCGTG ATCCAGACA ACCCTCTCG 360
CTGGGCTCC TGTCCGCTTC CGAGCATGT TTGGGATG TTGGGATG CCTTACAGC GTTTCAGAA 420
AGGAGATCG CTAGGAGGG TTGGGAGCT GACAGAGCC ACATACGAG ATAGAGGTA 480
CAGAGATAG TACTCTCTC AGTTGCTCG TGGAGTCAA TATGCTTAT TTCTATGAGA 540
GATGAGAT AGCTTCCAG TGTGCTATAC AGCGCGACA CAGAGAGCG CTTACAGCG 600
GATTCAGTG AGTTTCCC AGGAGAGCT CGGCGAGAC AAGATGTC GGAAGCTTT 660
GAGTTTAC CTGAGTCC TCCCTAGAG TTCCAGAGG AAGAGAGAG AGCGATTCG 720
ACTCCGAA AGTTTCGAA ACGATTTTG GTTTAGGAG AATGGATC AGAATTCGA 780
GAAACCGCA GACTCTTAC TTGAGTTTG TAGTGATTC GAGTGAAG AGCAATGGA 840
TTTCTCTCG TTGATGAGA TCCCTACTT GGAAGTATCA GAGGAGAG AATTTAGTG 900
TTTGTGGGCC CTAGATFAT AGCAAGAGC GTTTGACCG ATCACAGCA GAGTGAGAA 960
GCCACTGCG ASATGAGCG GCATCTTCA CACTGTGAG ACCAGAGCG ACCCTGCTAT 1020
CGCGAGCTG GCAAAAGCT AGGGGATGT GTTTGCGACT GATGCGATC TGGCGAGCT 1080
GATGAGCTT ACCGCTCAG TGTATTCTG GGAATTTTC GTTCAGAGAG TTGCTGCTAA 1140
ACTCTCTTT GAGAGAGAG ACGACTTGA CTTTGACCTC CTGAGCTGA GTGAGACTC 1200
CAATGAGCC CTCGAGTG AGGTATTTTC CTTCAATTC CCGCGAGCC TGGCCTGGA 1260
GCGAGCTAC ATACAGCA ATTTCTCCA GCACTGCTTG AGATGCGGA AGCAAGATA 1320
CACTTCCC AACCGAGCC GTTTGTGGA GAGGAGATG GATAGATG AATTTGCTTC 1380
TGTTCGTAC CTTTAGCCA GTGGAGCT TTGAGATG ATTTGACTTA TTGCTGCTTG 1440
TGAGAGAT GCGCTGTA CTGAGCCA CGGGAGTG TCTTCTTCA ACATCAGAC 1500
ACTCAATAG TGGAGTCA GCGACTTAA TGGGTTTAC TGGCTGAGA AGCTGAGCT 1560
TGAGAGGG CTTGCTATT CAGAGAGT GAGAGAGC AGCTAGAGT TGGCGCTTG 1620
GACTGCTGT CTTTCTCTG CTGAGTGA GTAGCTGAG CTTGTTATG TTGCTGCTGA 1680

60

CGAGTCAA GACTCTCAC GCGAGCTAT CTTAGGAGC CAGAGTTCA AGCTAATGA 1740
GTTTCCAGC CAGATCAC TCGAGTGA GATGCTCG GCGATTTTAC GCTGCTCAT 1800
TCACTATC ATGAGCTCG AGAGGGGAA ATACTATC CTCAGAGC CCAAGAGCA 1860
GCTATCCT GTCTACAGC TCCCTGATG CAGCTTACG TGTATGAG ATGAGAGGA 1920
AGAGAGAG GAGAGAGG AAGAGAGCA GAGAGAACT TAAAGAGT ATGTGAGCT 1980
GGAATTTTC CTTCAGCA GACTACAGG GCTTTGATG CTTATGAAA TGTGCTTCA 2040
ACTGCTCT TCAATTTAG CAGATGAAT AATATATA TCTGTTAT CTTAAGAAA 2100
AAAAAAAA AAAAAAAAAA 2120

(2) INFORMATION FOR SEQ ID NO: 161:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

30 GGAAGTCAA GTCTTTCCAG ACCAGGACA ACCAGGCTT TCTATATGA GCTCAGACA 60
CTTCAGCT CAGCTTTTAC TCGAGAGAC AGAAGCAAA GTTCTACATC AATGAGATG 120
CGAGATCG GCGTTTTC ATTAGAGCA ACTTCTTCA GCGGCGCC AGGCTCTGC 180
AAGTACAAA GTGAGAGAG CTGTACTGA CCGACTGCT GCGATGCTT ACTTGCATG 240
GTTTGGCTT TCACAGGAC AATACAGGT TCTTGTATT ACCGAGCTG GAGAGAGCC 300
TTGAGTGGC CTTGAGATC AGCGAGAGC ATGTCTGTC CAGAGAGTC TGTGCTCAG 360
GTGCTGTC GCTGCTGA TCGCTGAG TTCTGCTAG AGATGAGTA TTTTATGGA 420
AATGAGAG CTGAATAT GTTTGTGAT CAGAGAGC AGATCAGGT GACTTTTGA 480
GCTATGCT TCGATTCG CTATGCGCA ATGCGAAG ACCTGCGTA CTTGAGAGC 540
ACGAGAGC CTCAGAGGG GAGAGTTGAG TTGATGAGA TGGAGTGA CAGAGAGAG 600
GCGGCTTCC GCGCGGCA CTTCCAGAG CTTGCTACT GCACTGAGA GTGCTCTAC 660
GGTTTTCG CATGAGAAA TTGCTTCC AMATGAGG ACHTATGAA GCAAGAGAG 720
AAGTTTGTG ATAGCGCG GCGCTTCTG GAGCGCTCG GTGAGTAT CAGGCGCTCA 780
GAGAGCTCG AGAGTATCT GAGGTGATG ATGCGCTCA CTTATAGGA GAGAGCGCC 840
TACCGATCG TCGAGAGCA CTTAGAGCT TTGCTGAGG ATCTGCTGT GTCTCTCAT 900

60

(2) INFORMATION FOR SEQ ID NO: 162:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1003 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

5 GGCACAGAT GAGGGGAC CAGTCTCT AGGGAGGCT GGTGTGTGT CCGCTAGTA 60
15 TCAGCCCTC TTACTGTACT CTCGGGAT GTTAACTTT CTATTTTCA GCTGTGAC 120
CTGTCTAGGC AGCTGTGCTT CCCATTGG CCTGTGTGT CACACAGAG GTGGCTGCC 180
20 CCGAGGGCA CCGCTCTCTT CTGTACTC TTCTCTTAC AGTGACTTGG GCTTGAAT 240
GCCAGCAGC CTTCCTTTTA GCTTCACAC CAGAGAGA GTTGTACATG ACCTCCGCG 300
CCCTCAGCA AGCTGTGGA CAGAGGGAT GTGTGAGAG CAGGTTCTCT GTGGCCCTCT 360
25 CCGGCTGTT TTCCACTAGT CACTACTGTC TTCTCTTGT ACTATATCA TCATATTTCT 420
TCCCTTGGT GTGGGAGTG GAGAGGCTGC TGGGTGAG CTCGACTTC CCACTAGTT 480
GGGAGAGAG GATATACAT GAGCACTGTT CTGCTCAGAG CTCCTGATCT ACCGACGCC 540
30 CTAGATCCA GAGTGGGTC AAGCTGCTAT GAACACGCG CCGTGGAGCA AGCTGGGAA 600
TGGCTGAGG TGGAGGAAA CTTGAATCTC TCTTTCCGTC TCCCTCTGCG AGCTTACTG 660
35 GAACTGTATC CTATTAGAT CTATTGAGT TGTTTCCGT CTCGTTGGG CAGAGAGCA 720
AGGAGAGCG AGGCTCTAG AGGCGAGCC CTTCCTTGT CTCGTGGGTA AATGAGCTG 780
40 ACGTAGATA AATGAGAGA CAAAGCCCT CTGATTTTAA ATTTCTATA AATGTTAGA 840
GTATATATAT AATATATAT ATTTCTTAA ATTTTTCAGT CTATTGATG TCTAAAAATC 900
CATTTCTCT CCGCTGAGC CTGATGAGA CAGTGAAGA AAGCTGTGTT TCATTTAAG 960
45 ATGTTATTA ATGATTTGA ACTTGAATA AAAAAAAAA AAA 1003

(2) INFORMATION FOR SEQ ID NO: 163:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2196 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

5 ANGAGGCGC AAGAGGATG CGATTCTTAC ACGAGATGG GAGAGATAC CAGCGACTTG 60

GGAAACATC AGCATATGCA TCAGCGAGAT GACTCTATG CTCGACAGAT GGAACGAGA 120
5 ATGAGGACA AACTGAAAC AGCTTTTAA ATTTCTATG ACAAATGAGA GGCCTTAAT 180
AAGAGAAC TCGAATTTGA AGTGGCTTTT AGGACTTGG GATTTAAGCG AGCTGCTAT 240
AGAGTAACT GCTTCCTTCA CCGCACTAGT AGTGGCTGG TAAATGCTAC GGAATGGCA 300
10 CTTTTTGGG TCAGATTGGA TGAATGAGAG CTGATCCACT TTAGCGGGT CCAATTTTAC 360
CTGAGAACT TTGATATGTT ATCTCTCTAC AAGGACTTCA GCAAGAAAGT GACCTGATC 420
15 AACCCATTC CTGTAGCTTC TCTTGACGCC ATCAGGAAT GTTGAATTC CTCGACCTG 480
AATACACAG AAGGATGAGA GTCCCTGAC TGAATTAAA TCATGAAAGC CATTTGTAT 540
GACCTGAGG GCTTCTTCA ACAAGTGGC TGGCTTTTCC TCGAGCTCA GGTGAGCGG 600
20 AGTATGCTG AAGAGGGGA TTGAGATCT GAAATTTAG ATGAGACTTT TAACTCTTCA 660
CAGATGACT ATGAGAGA AAGAGAGAC AGTGAAGAG ATTATTCAT AAGAGCAGA 720
GACTGAACT ATTCTAGCA GTCTTGGT AGTGAAGAG AGATGAGAA GAAATGGAT 780
GACTGAGG AAGAGAGCG AAGAGAGAC CAGAAAGTC GTTACGAGA AAGAGAGA 840
900 CAAATGCGA GTATGAGCG GAGAGAGAG GATCTGTGC ACATTTGGG CCGTGGCTCT
30 AAGCTGCTT CAGAGCAG CTCCTGACC CCGAGAAA AAGAGAGTA ACTTCTGAC 960
TTTTGGCTG AGTCTCATTC TTCTCTGAG CAACTCTCA AATTTTACA TGAATGAA 1020
ACTATATTT TCTTTGCTT TCTATTTGAA GTTTTGGAT TTGTATTTAT GGTTTTGGG 1080
GGCAATTTGT GTGAGCAAT CTACTGGGG AATTCAGGC CACAGAGAC AGTGGCAAT 1140
GGCCCATTC AATGGCAG GAGAGAGTG TTCTTGAAG CAGAGAGAG CTCCTGCTGT 1200
1260 TAAATATAT TTTTCTATC TTCTCTCTTC CTGTGACTT CTCGACAGC ATTTGATGCT
1320 TCTGACTCT TATTTGCTGT CTCAAAGTG TATTTCCAG ACGATGATAC AAGGTGACC
1380 TTATTTACC GTATCATGTT TCTTGACAG CACATTCAT CTCCTCACT ACCCTACTGC
1440 CATGACTTC CGCATCTC TAAATTTTAT CTTTGCATA CTCAGGTTTC TCGGAATTT 1500
GCTATGCTT GTGATTAAC ATACAGCTTG ACCCATGAG GCAATTTGG CTGTGCTCT 1560
50 CTTCTGAGT TTCTGCTTT CTCGCTGCT GCAATTTCT AGGTATATCT GCTGCTTGG 1620
AAGACTTAA AAGAGATAT ACTGCTGCG TGGTATTTT TCTCTATCA ATGACACT
1680 GTTACATGA TAGAGGCA AATTCCTCT GTCTCTTTT TTTTCTATA TATCTTANGA
1740 AATATTTCA GGTGTGCTT CAGTACCG TTCTATGAA ATGTAGAGA AGGCTAAG
1800 GATCAACT TTGATCTGG AAGGAGAG TCATGCTTG GCGTAGAAT ACCGTATGA
1860 GAAGAGAGA GAGAGGGA GCGATATCT ACGAGCAAT CTTCTGCGA CTGCTGCTC

5 TTTATTTTAACT TTTTCTCTTTC ATTTCTCTCTT ATTTATTCACA GTTTCCTCTTG AACAGCTTTT 1920
CAGATTTTGG GGGAGTTTAT TTTCCTATTC TCCCTTTCTTG GTTCTCTGCA CCGACTCTTC 1980
CCACTCCAGT TCTTCTCTCTG CTCTCTGACT TTATAGAGAG AAGCGCGGAG GGTCTCTCGA 2040
TTTTATGTTT GTTGTGTTT TCTCTTTAGC AGTAGGACTT GATATTTTCA ATTTTGGAG 2100
10 AACTAAAGCA TCAATTAAGT GGGTTTTT TGTGTCTTCT TTTTGTAA AAAAAAAAAA 2160
AAAAAAAAA AAAAAAAAAA AAAAAA 2196

(2) INFORMATION FOR SEQ ID NO: 164:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1945 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

5 GCACAGATC GCGCGAGCG ACAGGAGAG GAGGAGGG GGTCTGCGCG GCGCGCTTAC 60
CCAGAAAGCA CCGAGAGCCA GACAGGAGTG GGTCTGCTCC GAGCGCAGCC CCGAGCGAGC 120
CCCCCGCCCG CCGCGGAGCG AGCGCTTTC CAGCGCAGCC GACTCTCTAG AGGAGGCGAG 180
CGCGGAAGCG AGCTCAAGCC TCACACAGCG CCGTGCCTCC AGCGCGCCCA CTCCAGGCT 240
CTCTCGGACT CCGCGGCTTC TCTCTGAGT CTCTGAGGAG ACAGGCTCTG CAGAGCCAT 300
GGAGTTGCTG CTGTCTCTTC TCTGAGGCT GCTGCGCCCG ATGCTCTCTG CCACTGCGAC 360
TGAAAGAGAG AAGGAATGCG ACCCTTTTCA TTATGATTTAC CAGAGCCCTCA GATTTGCGCG 420
ACTGTGCTTC GCTGTGCTTC TCTCTCTGCT TGGGATCTCT CTATCTCTAA GTGCGAGGTC 480
CAAGTCCAGT TTTATCTAGA AGCGCTCGGC CCGAGGAGT GAGGAGGCC AGCTGAGAA 540
CTCTATACCC CCGATTCGA CAGAGCGCCA GAAGAGAGAG AACTGAGTG CAGCCATCAG 600
GTGGAAGCTT CTGAACTCTG AGCGCTCG CTGAACTTTT TTGAACTTTT GATGCTCTCT 660
AAGAAAGCG CCGACTCTAG CAGAGCGCTT TTCCCGAGGA GAGCGCAGA ACTTGTCTGT 720
CCCCAGCTCT ATCCCTCTCA ACAGCTTCC TCCAGCTCAT GATGCACTTA ACAGCTCTCT 780
CCCCACTGCA CCGTCTGCTC CTGCGGCTCT CCGCTGCTCT CCGTCTGCTCT GTTGTCTGT 840
55 GTGACTCTGT GTTGTCTGTA ACTGTGCTCT TTGTGCTTAC TTGTGCTTAC ATGCTGCTGT 900
GTTGTGTTAGT GAACTGTGTA CTGCTTCTTC CAGCGAGGCG CTGAGCGACA TCGGCTCTG 960
CTCTCTCTCT CCGCGTGGC CCGCTATCAC CTCTCTCTTC TAGGAGGCTG CTGTCTCTCC 1020

5 GAGACAGCC CCGTCCCTG ATTTAGGAT GGTAGAGGTA AGAGCAGCGG CAGTGTCTTT 1080
CAGTGTCTT GGGACTCGG AAGGTTTCCA GCACTTTCTC ATCATCTCTC ATGACTCTCT 1140
TTCACTCTTT TAACAAAGAC CTCTCTCTT TATCCAGCTT GATCCAGTCT TGAAGTCTTC 1200
TTAGAGCTG GAGATGAAA GCAAGAGCT GGTAGAGCCA GGTGTGACTT CAGCGAGGCT 1260
ATGCGCTTCC GTGTTATTT TCTTCCAGG GGTTCCTCAG AGGATTCGCC ATCTGCGCG 1320
10 CCGCTTACA GAGCGCGCG GATTCAGG CCGAGGCTT CTAGTCTGCC CTTGCGGAT 1380
GTGTCCCTG CATTCTCTT CAGCATTAC TCGATGCGGT CTGGAGCTCT ACCCTTTCCA 1440
15 ACCTTCCCTG CTCTCTGAGC TTCAATCTAC AGCGCAGCTC ATCCAGATCC AGACTACAT 1500
CCCTCATTT GGTCTCTCTG CAGCGATTAG TTGAGAGACT CTGTCTCTCT TCGCGCGAGC 1560
ACAGCGGAT GATGAGAGG AGAGCAGAG GCTTGTCTCT TCTGCTTAGG TCCCTTTAGA 1620
20 TCGCGAGAG AGCGACTCC CCGATCTTT CTCTCTCTG TCTGTGCTCA GAGCGGTAG 1680
CAGGTGCTT TCGAGCTCA GCGAGCTCC TCGAGGCTT GCGAGAGTG AGAGTTTGA 1740
25 GGTCTATAGC AGATTCGAA CTGACCGAG ATCCGCGCC TCTCTCTCT TGTGTCTCG 1800
CGAAAGCAA CCAAGCTTG GGTCTGAGC CATTCTCTT CTCTGTATCG TGAATCTATC 1860
TCAGAGCAA CAGAAAGAG GATTAATTA TCTTTTCTT CTGATGAAA AAAAAAAAAA 1920
30 AAAAAAAAAA CTGCA 1945

(2) INFORMATION FOR SEQ ID NO: 165:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2933 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

45 GGTGTGAGCC AGCGTCTCG CAGCGCTCTT TTGATCTCTT GCTGCGCGCTG CCGCTCTCGG 60
GATCAGAGC CAGTGTATAC GCGCGCGCA CCGCTTGTGT GCGCTTAGAG GAAGCGAA 120
50 GAGGCGCGCG TCGGTCTCTT GCGCGAGCT CCGCTCTCTT CTGAGAGAG CAGCGCGCG 180
CCAGTCTGT CCGTCTGAC CCGTCTGAG CCGTCTGAG CCGCGCGCAG AGCGCGCGC 240
CGGAGAGCG CCGCGCGCAG GCTCTCTGAG CCGATTCAAA AGTCTATGAC CTATCAGAG 300
55 CATTCTCAA ACAGAGAGC CCGCTCTATC GGTATTTCT TGAATGAGC ATCATGAGG 360
ACAGAGTGT GGTGTCTCTG TCGAGCGCA AGATGATCA ATTCAGGTTG TTGCGAGGTT 420
60 ACAGAGTGT CCGTAAAGA AGAGAGAG CAGAGAGCTT TTGATCTCT CTCTCTGATG 480

5 ATACTTCTTC TCATGAGAG ATTGCGATCA ATAGAGTTGT TCGGATATAC CTTCGTGTAC 540
GCTTAGGGGA TGTATACAG ATCCAGCAT GCGCTGATGT GAGTATCGGC AAGCATATCC 600
ATGTGCTGCC CATTGATGAC ACGTGGAGG GCHTACTGCG TATCTCTCTC GAGGTATACC 660
TTAAAGCTTA CTTCCTGAAA GGTATCTGAC CCAATCTGAA AGGAGACATT TTCTTGTCTC 720
10 GTGTGGGAT GGTCTCTGTC GAGTCTAAGG TGTGTGAAC AGTCTCTAGC CTTATATCA 780
TTTGTGCTCC AGACACAGTC ATCCACTGCG AAGGGAGGCC TATCAACGA GAGCATGAG 840
15 AAGCTCTCTT GATATGATTA GGTATGATG ACHTGTGTG CTGCGAGAG CAGCTACTTC 900
AGATTAAGGA GATGTGTGAA CTGCGCTGCA GACATCTCTC CTTCTTTAAG GCAATGTGTG 960
TGAGGCTCTC TAGAGGATTC CTGCTGTAGC GACTCTCTG AAGAGGAGG ACCCTGATTTG 1020
CTCGAGCTCT AGCAATAGAG ACTGAGGCT TCTTCTTCTT GATCAATGCT CCTGAGATCA 1080
20 TGAGCAATTT GGTCTGTGAG TGTAGAGCA ACTTCTTAA AGCCTTTGAG GAGCTGTAGA 1140
AGATGCTCTC TCCATCATC TTCAATGATG AGCTAGATGC CATGCTCTCC AAGAGAGAGA 1200
25 AAGCTCATG CAGGTGTGAG GCGGCTATG TATCAATTT GTTGACCTTC ATGTATGGCC 1260
TAAAGCAGAG GGCACATGTC ATTATATG CAGCAACCA CAGACCCAGC AGCATATGAC 1320
30 CAGCTCTAG GCAATTTGCT GCGTTGACA GCGAGGTAGA TATTTGAAAT CTTGATGCTA 1380
CAGAGCCTTT AGAGATTTCTT CAGATTCATA CCAAGAACAT GAGCTGTGCA GATGATGTG 1440
1500 ACCTGGAACA GTAGCAATG AGATCTAGG GCHTGTGGCT GCTGACTTAG CAGCCTGTG 1500
35 CTGAGAGCT GCTCTGCHAG CCAATCTGAA GAGATGTGAT CTCAATGACC TAGAGATGTA 1560
GACATTTGAT GCGAGGTGTA TGAATCTCTT AGCATTTACT ATGGAATGACT TCCGZTGGCC 1620
40 CTTGAGCCAG AGTAAGCCAT CAGACTGTG GCAAAACCTG GTAGAGGTGC CAGAGGTATC 1680
CTGAGAGAC ATCGGGGCC TAGAGATGT CAAAGCTGAG CTAGAGAGAC TGTCTCTGATTA 1740
45 TCTCTGTGAG CAGCCAGACA AATCTGTGAA GTTTGTGATG ACACTTTGCA AGGAGATTTCT 1800
GTTCATATGA CTTCTGTGCT GTGGAAGAC TTTGTGTGGC AAGCCATTTG CTATGATATG 1860
CGAGGCCAG TTCAATCTCA TGAAGGTGTC TGAAGTCTCT ACHTGTGTGT TTGGGGATTC 1920
50 TGAAGCCAT GTCAAGAAA TCTTTGACA GCGCGCCGAA GCTGCGCCCT GTGTGTATTT 1980
CTTTGATGAG CTGATTTGCA TTCCAGAGC TGTGTGAGGT AACHTTGAG ATGTGTGTG 2040
GCTCTGTAC GATGTATCA ACHTGTCTT GAGAGATG GATGTGATGT CCAAGAAAA 2100
55 AATGTGTTC ATCATGTG CTAACAGG GCTGTGATC AATTGATCTG CCACTCTGAG 2160
ACCTGGCCTT CTGTATGAG TATATATAT CCAATTTCTT GATGAGAT GTGCTGTGTC 2220
60 CATCTCTAG GCTAAGCTGC GCAATCTGC AGTTGCGAG GATGTGACT TGAATTTCTT 2280

5 GCTTAAATG ACTAATGCT TCTGTGAGC TGACTGACA GAGATTTGTC AGCTGCTTG 2140
CAAGTGCC ATCTGTGAT CATTGAGAG TGAATTTAG CAGAGAGAG AAGGCGAGAC 2400
AAGCAATCA GCAATGAGG TAGAGAGCA TGAATCAGT CCTGAGATCC GTTCAGATCA 2460
CTTTGAGAA GCAATGCT TTGCGCGCC TTCTGTGAT GACATATCA TTCTGAATTA 2520
10 TGAATTTT GCGCAGACC TTACAGAGG TCGGGCTTT GCGACTTCA GATTCCTTC 2580
AGGAGACAG GGTGAGCTG GCGCAGTCA GCGAGTGA GCGGZGAGG GTGZGAGT 2640
15 ATACAGAAA GACATGATG ATGACTGTTA TGCTAATG GTGTGTGCA GCTGTGATG 2700
AGCTGCTG CTTGAGCTT GTTCTGTG GGTGCGGCG GTTGTGCGAG AAGAGAGCA 2760
GCGTGGCC CAGAGGCTGC TCAATCTTC ATCTGAAAC GTTCACTAC AGTCTGACTC 2820
20 TCGAGAGGG GTTCTGTG CAAATATCA AAGCAAGGC GATTAATTA AAGCATTTT 2880
CATTTGTAA AAAAAAAAAA CCGCGCGCG GCGCGAACA TTT 2933

(2) INFORMATION FOR SEQ ID NO: 166:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2243 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

5 TCGAGAGCC GCGCGCGCG GCGCTCTG CAGAGAGCG CTTCTGTGAC GGTGTATAC 60
GATGCGCAGA AGTGGCTTG GCGTGGGAT CACATAGCT TTCTAGCTA CCGTATCAG 120
GCGTTTCTC GTGTATATG GTTCTATCA GTATACATCC CCAATTTCC TCTATATCG 180
TTCTGCTC CTTTATAT TTTTCTGAG AGCGTCAAG GTGAGACA TAGGAGACA 240
45 GTTAGCTATG GGTGTGCTG AAGGCGCCA TAGTATATCA GTCTTCAAA CCAAGATTC 300
TGAAGCAG GAGATTTTG GAGAAATTC TGAATGTGA TTATGAAA GATTTCTTT 360
50 TTCTTATAT AATCTATTTA GATCGGCTG ACTGTACAA TGAATCTG AAAAACTCT 420
TCACTAGTC TAGATAGGG AGTGTGAAA TGAATCTTA CCGTGAATC TTCTCTGAC 480
TGGCGGACT GCGCGCTGC TGTGCGCTG AGCATCTGC CCAAGTACG TGGZTTCAG 540
55 CAGTGGCAG CTTCCGAGT ATTCGATTC ATTATCTCA TTAAAGAG TTGCGATTT 600
TCAAGCTT GACTATGAC TCAATTACA ACCCGAGTT TTGTGTGAT GCGCAAGGA 660
60 GGTATGTGA TGTGCTTAA GAACTGAAA GTATGTGTTA ATAGAAATTA TATTTATCA 720

5 AAGATATTTT AAAATAGGG CTGTGTTTAA AAAAAAACC AAAACAGGAA AACAGCGCT 780
GATTATAGAG AGGTACAGCT CTAAATGGGG TCGGGGGCTG GCGAGCGTTC ACCTGACGC 840
TCTCTGCTGC TCGATGGGG TGTATACAGT GTACACAGTG TGTGTATCAC CAGTGGGTCA 900
ACTGCTGTGC ATTCCTCCG TGGCAATTTG TGTAGACAACT CTACTAGAC AAAAGCGAAT 960
GAAAGTCTTT GGTTCGCACA CTGGCAATTA TTGGAAATTT CAGCTCAGTT TATGAGTTT 1020
ATTTCGAAT CCAATGCACT CTAAAGATCA ATACCTGTCT GCAATGTAT TCAATCTTAG 1080
TGAGCCAAA TTGTTGTTT GTTACTACAG AATAGAGATG ACTGTTTTTT GCGACAGCC 1140
TATGCAATTT GCAATCTGTG ATTCCTTGT AAAAAGGAGA GTGCATATG CACTGCATTA 1200
AAGTGTGTCT GTTCTATGTC ATGATATG TGAGGACAA TGTATATAT TTATGGCAAT 1260
GACATACCA GACCTAATTT GCAATGATG GGTCTTAAAC TTCAAGTCCA ATGTATATCA 1320
AAACCAATCT GAGCTGTGA TCTCTTAAT ATTTATTTT TTTAAGCTGT GAGATGTG 1380
AGAGAGGTT CTCCATCAT TTCAAGTCTG CCTGAGGAA ACTCGCAAT GATTTCTTTC 1440
AGTGTGAG TTCTTTGT GTTACAGCTT CAGTGACC CTCAAGCTTC GAATATGCC 1500
AGTTTGTG GTTTGTGAT TTATCTAT ATATTACCT TTGTATAT TTGCATTTAC 1560
ATGTGTTG TTGTTTAA ATCTGTGAA AGTGGTGA TTAAAGACT CTTTAAAT 1620
GGAAGCAAC AGTCAGCA ATGAGAGCTT AGAGAGCTT GCTGTGAC GCTGTCTTT 1680
GTGTTGTT TTGTATGA AGATCTTG CTGGGTTTT TTGTTTGT TTGAGGAAA 1740
TGTCTTGAG TAAATTTAA GTTCTGAG TTAAATTTT TTACAGCAAT TTGTTT 1800
AAAAATAG GATCTCTG ACTTTTGA TGACCTCTT ATATTTTC TGAATGAA 1860
AACATTA CAAGAAAT TTCAATGAA GATGTGACA TTATATGAA AACAGAGT 1920
TATATGTA AGCAGGAG TGAATCTTTA AACAGACT GTGAGGAC ACAAATAG 1980
TCTTCTCT CGAAGCCGA AGTAAATCTA TATCTGTAG AATATATGA GCGAAAGA 2040
TGTAAATTT AGATTTTT TTCCATAG TTATGAGAAA TATATGACC AAGTGATTT 2100
GAGTTTAA AATGTAAA TATATGAA AACATTTGA CTCTACCA TTGAGATC 2160
TATGAGTT CAGATCTA CTAAATTTT ACATGTGT TCTTTTTGA TCTATTTT 2220
ACTTTATTA AGGTTGAA ACC 2243

(2) INFORMATION FOR SEQ ID NO: 167:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1816 base pairs
(B) TYPE: nucleic acid(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

5 GGTGCGAGC TTATATTTT CCTTACAGG GCGCTATTA GCGGAACCT TCCCGAAT 60
TTGCGGTGA CCGAGCGTC CCGCAGGCT AGGAGAGAA GTTGTATGTC CAGAGGTGA 120
GCGAGAGGC GCGATTTCT GCGCGGTGAG GCGCGAGCT AGTGTACAC GAGGCGGAA 180
GCAAGGTGTC GTGGGCGGA GAAAGGGGCT GCGCGCAGA GAGAGGAA ACCCTTCCA 240
GAAACAGCA ACAGCTGAG CTGCTGTAC AGAGGGAC AGATGCGCG GCGCGAAGG 300
GAGGCTCTG GTAGAGACC AACTGGGCT CCGCGCGCT CTGCTGTGA CCAATGGCTT 360
GCGCGAGCT TCGGGAGCG CTTCGCTGA AGCATTTGAC TCGTCTTGG GTATAGGCG 420
GTCTTGCCAC CCGCGCTGTC AGTTGAGCTA CCGCTTGAC ACTTACCTTA AGGAGAGAA 480
GTGTAGCCA TGTACAGAG GTTCCAGCT GTTTTCAAT TTTCATTTT TGTATTTG TGAATGAG 540
AATGACTTA AATGACTTA AATGAGATG TGAATCTGA TGTACAGAG CATATTTCCA 600
ATCTGATGAG CAAATGCTT GCGATCTTG GTGCGAGAT CAGCTGCAAT TCGCTGACT 660
GAGAGAGAA CAACTTATCT CCGTATGCC AAAAATGAC CTACTCTTTC CTCTAACTCT 720
GCTGAGCTA TTCTGAGTG AGATGAGTA CTTGCGAG AGCTTCTTAA CTTCTCTATG 780
GACTTTTAT CTTCAGCGC ATGAGGAAA AATGATTTA TTCTGTCTA AGCCAGAAA 840
TCCAGGATAC GCAACAGAT TGGAGCGAG AGCGCTTACA AATTTGAGAG RAACTCTCT 900
AGCAAAATG TCTTCAATCT GCGAATGAG AAATTCAG GCGACAGCA ATTCTCTCA 960
AGATGAGAA AGTGTAGCT TTTTAAGATG CTTCTCTCTT AACTGTGCTT GATTTTAC 1020
TACAACTCT GTCTCTCTG TGAATGATTT GCTTTGAT TTGTTGTGAA CTTGTTGCTA 1080
CAGCTGTTG GACCAATAT AGTTTCCCT TGAGAGCTG AGTATGATG GTGACTTGA 1140
GTTATGATAT GAACAAAGC TAAACAGTA TCCAGCTTCT TCTCTGTG TTTGATGATC 1200
TAAACTGAA GATCTGAG AGCGAGGCT TCTACTTACA AAGTGTATC TTGCTGATTC 1260
TGAATTTAA GCAATTTTCT TTAAAGAAC AGTGTATTA GACATCTTAA ATTCCACTCC 1320
TCAATGAGCT TTAAATGAG TTTCATGGA TATAGGCTT AGAAATCAC TATATAATCC 1380
AATTAAGTT ACTCAATCT GTGAAAAA AAAAAAAA AAAAAAAC TCGAGGCGGG 1440
GCGCTTACC AATGCGGCT ATATGATGAT GTATTTAT TTACTATTA TCTGTAGCTA 1500
TTTGTTTTT KCTTTGCTT ATGTTTTT TCCCTTTCTT MAGCTATAG CTGATCTATG 1560
CTGCTCTCA CTTCTGCA TGTACTGTC AGTATCTTA GTTACAGCTA GTTACAGCTA 1620
GTAGAAATGA TGTCTCTCT CAGGATGCG CCAAAATCT GTAAATTTGA ATTATAGAG 1680

AAATGACCTT TATGACACT AGATTTCAG GACTGAAAT CATTAAGATT TTATTGCAAT 1740
AATTATGTC TCAAAAAA AAAAAAAA AMMRARASK RRMACTGCA GGGGGGGCC 1800
GGTACCAAT TCGCCG 1816

10

(2) INFORMATION FOR SEQ ID NO: 168:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 945 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

AGAAACGTT GATGGCACTG AGAAACGCA GTTAAAGCT CTTTGGAGCT TCTGAGACT 60
CACTGGAAC CAGGGGGAC AGTTGGGAC ACCATCACT TCTCCGAC AGAGAAACC 120
GAACCCACA ACCAGGGCA GATAGCTG AGAAACATC TACAGGCA AATCAAGTT 180
ATTGGACTA TCGAATGTT GTTGGGATG ATGATATGA GTTGGGAT CATTTTGGCA 240
TCTGCTTCT TCTCTCCAA TTTTACCAA GTGACTCTA CACTGTTGA CTTCTGCTAC 300
CAATGATG GACCTTTT TTTTATATC TCTGCTCTC TATCATATC CACAGAGAA 360
AGATTACCA AGCTTTGCT GATAGGAC GTGTTGGA GCATCTGAG TCTCTGCT 420
GGCTGGTGG GTTTATAT CTTGCTGTC AAACAGGCA CTTTAAATC TCCCTCATG 480
CACTGTGAGT TCGACAAA TATATACA ACAGAGTT ATGTTCTTA CTTTATCAT 540
GATTCACTT ATACAGGCA CTGATACA GCGAAGCA GTCTGGCTG AMCTCTCT 600
CTGATCTGA TTTCACTCT GCTGAAATC TGGCTAGCT TGTCACTGC TGTCTGGCG 660
TGGAAACGG CTTACTCTGA CTCTCTGG AGTGTACTTT TCTGGCTCA GATTTACAT 720
GCTAATCTG GCATCTCTC AATATGACT CATGACTGT GATATGACA ACTATGACT 780
TTTAAAGAA AAAGGAGAA ATATATACA GAAGTTGAT TCTATATTA ATATGAGAA 840
GTTACCAAT ATAGAAAGC AAAGTTGAG TTTCTTAAT GTAGCTTTT AAAGTATCA 900
ACATAAAA AACCATAT TTAATCTA TTTAAAGTA ATGTC 945

55

(2) INFORMATION FOR SEQ ID NO: 169:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 902 base pairs
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

GGGAGACCA CAGAGGAT GAGGAGACC AGGCTCTGG GGTCTCTGT GATCTCTTT 60
GTCTCAGAC TCGAGCTGC AACTAATTA ACTGAGAAA AGTATGAACT GAAGAGGGG 120
CAGACCTCG ATGTGAATG TGACTACAG CTAGAGAGT TTCCGACAG CAGAAAGCT 180
TGGCAGATA TANGGAGGG AGAGTGGCC AGAGCCCTGG CATGCCAGG GAGGCTTTCA 240
AGAAATGCC ATCCATCCA AGTGGGAGG ATCACTACT AGACTACCA TCAATCATGT 300
TTACTGCGG TCGATGCT CAACCTTCA GTGGAGATT CTGACTCTA TCACTGTGTG 360
ATCTACGAC CTCCAGGGA GCTTCAGAT GTTTCGATC GCATCCGCTT GCTGTGACC 420
AAGGTTTTT CAGGGACCC TGGCTCCAT GAGATTCTA CCCAGATGT GTATAGATT 480
CCTCTACCA CCACTAAGG CTTGTGCCA CTCTATACA GCTCCAGAC TGTGACCCA 540
GCTCCACCA ACTCACTGC GATGTCTGC ACTCTGACT CTGAATCAA CCTTACAAAT 600
GTGACATA TCATCAGGT TCCGTTATC AGATTTTCA TTCTCTGCG TGTGTGATTC 660
CTGATAGA GCTGTCTT CTCTGTCTG TTTGCTGTA GCTGAGCTC ATTCTTACC 720
TAGGCCCCG AACCGAGG AATGTCTCT GACTTCAGC CAGATCCATC TGGGATGT 780
GCGAAGGAG GAGGAGGAG GTAAAGGCA GGAATTAAT AACTGATTA AATCTCTTA 840
TCACTGCTA AAAAAAAA AAAAAAACH GAACTCTGG TTTTCACTC CATCAGCTC 900
TT 902

(2) INFORMATION FOR SEQ ID NO: 170:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1883 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

AGAAACGAC TGAAGACA CATTTTCTA CATACAGCTG GCGAGTAGC TGAAGACTG 60
GCACTGGCA CCACTACTAG GTTGAGAG AGTTAGGAA ACCAGAGGC CAATGGATC 120
TCTCGGAAA CCTGAACCT GTCTCTGGG CTCTCTTAC AGTTCTGAG TTGAAATTC 180
TTTTATGCC TAGCATCTG TTGATTTTA ACCCGAGG CAGCATCTC ATAGACTAT 240
GTTTACTCT GTTTTACTT TGTTTTAAT CTCTTAGA CCGAGTGGC TCTCTCTTT 300

5 TCCCTCTTGG TGGTACGCTC TGGCCATCTG GAGCTCTAMT CCCGAGCTTT CCCACTTTCA
GCACTGCTTTT GCTCTCTTGG CTCTACCTC AATATAGCCC AGGAGTGGGC TTATAGTCTCC
10 AATATAGGAC ATTTCAGCTT TCTCTCGGG GATGGGGATT GGGATGGGCA GATCTGTTT
TGGTCTCGG GGTATTTCC AGTGGGTGTA AAGCGAGCC TGGGCTTTTC CCTCTCTTAT
CCCTGAGGCT GGTATAGAG GACTGTATCT ACACCTGTTTC TTCCCTACCT TCTCTTTTGT
15 TAGGAGGCCC TCATCTTAG TTCTCTAGA GAGTCTCTGG CTATAGAGTG TAGGAGGCT
GTCTTAGGTC GGGATTTGG AGGAAAGGCG TCGATAGGCG ATGATATAGG TATGAGATGG
GCTCTGAAA TCAGACAGAA ATGGCTTAG AAGCGGAGG GAGCATGCC TGTCTCTGAG
10 TATAGAGGTA TGGGAGGGAC CTCCTTAGCT TGGAAATGTA GATTTGAGG GGTATAGAC
AATAGAGTG CCTAGTTGAG GATGTGCCA AAGTTTGTGC CATCTTATC ATTAGTAGAT
20 TTATAGACC ACAGAGGTA ACAGAAAGCG GAATATGTTT ACTTTGAGTG GTTATTTT
TTGTTTAGG TGTGGCTTGG TACTGAGTA AGATGCTAT ATGCTGACA TTTTGGCTTT
1020 AAGTCTTAG GACTTTCCC ATTATAGCT AATGGGAGTA TACAGATGTC GAGTCTGCT
1080 TTTTGTGTTT TTTTATATAT TTTTGTGTTT TTGCTGTGTC TTATGAGAT TTTTACAGAT
GCAAGAGAT GAGGAGGTC GTCTTGGAC CCCATGTGTC CATGACTTAG CTCGATCACT
30 TATCAGCTAT GGTCAAGCTG GTTTTCACTG TATCTCTGTC TTTTCACTG TATTTGTTAT
TCAAAATCCA AGACACTGTC CCATGCGAC CCGTCACTAT TTGGGAGATT GGTAGTCTCT
35 TTTGATGTC ATAGTAGTGG GTTGCACTAT CATATACATA TCAAGTCTGC TTTTGTGTTT
TAAATGTATC TAAATGATTT CAGAGATGG GCTTTAGAAA TGTGTTTTTA GATTTAGCAA
GAGTCTCAA AAGAAATGTA GAGGATGCT TCCCTTGGCC TTGCATCTAC AATACAGAG
40 AGAGACTGTT CTGTTGTTAA ACTCTTTCAA AATTTCTGAT ATGATAGGCT ACTTGAGACC
CTTACACAGA ATGTCAATCT TTTTCTCTGT GTACATGTA AACTGTGTC ACCATTAGCA
45 TTGTTATGAG CTGTGACTGG TGTCTACT CTGGTTTTGG AAGATATAT TGGAAATTTT
TGTGTGTTTC TGTGAATA ACTGCGGCA AATATATAGT AACTGTGTTT TCTACTGCT
AATTTGAGAC CTTGTTATTA ACCGATAT TTCTGTGTTTC TTAAAGATTA TAAATAGTTG
50 TAAATTTGCA TGCATGATCG AAAAAATAA ACCGTATCT CTGTTAAAA AAAAAA
AAAAAAAAA AAAAAAAAA AAA

(2) INFORMATION FOR SEQ ID NO: 171:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2100 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

5 TACTTTTACA TTACTGCGCT TCMAAAAGTG CCTATTTCTGA GCACATATAA CATTATTCCT 60
10 TACATATGTA TGTACACAGG GTACCCAGAG TGTACTGTGT GCAGCTTTCA AACAATACC 120
ATCAGAAACA GTAGTGTCTG AGATAGGATA ACTTTGCCAA ATGTAGAAAT GTCACTCACT 180
TCCATATATC CTTCTTCTAG CGGTACCTCT GAAAGGGGCT GCMAAGACAT TCCCTTAGCA 240
TCCCTTGTCT ATACAGCTTC TTATATATTA TATCTTACTG GATGCTAGCA TATTCCTTAG 300
GTTTCTCTGA CTCTGCTTCA AGGGAATGTA AGTTTATGCG CATTCMAACA TTTAGGAAAA 360
AATAAGATCT TTAAAGAAAT TAAATAGACC GTAGTCTGTA TTAGATGTCG TGTCAATATCT 420
GTGTTCTATA AACTATAGAT CGTGGGTTTT AGATGTTTTAA AGTGTACAGA CATTCCTTCT 480
CCTTTTGTCT CTCAGCTTAA CATGAGAAA ATATAGAAAG TCTTGGCTGT GGGGATTGGA 540
AGCTCAGGGG GCGAATGTC CTTCGAGAT CTTTAGAGCA TTACTTTTGC TCTTAAATAT 600
ACTATGTTAT GTTATTTGAT GGTTTTGTCT TCCATAGTTC CATCACTGAC AATACTGTCA 660
ATATCTGTGA TGGAGCAGA CATTAGCTTA GATGTATGTA TCTTTACCA GAGTGTGCAA 720
TAGAGAGGCG TCCATGTAAA TAGAGAGGCG TAGACATGTC ATGATTTGTAG GAGAAGGTTT 780
GAAGGAGGCA CATGATTTCA AATAAGATCT TTTCTATAT TGTCTGTGCT CCATATTTCA 840
GCGAGATTAC ACTTGGCAG TGTGTTGCTT TCTTCTTAG TCAATTTGCT CCATATTTCA 900
TTGATATGTC CTCTGTTTTG GCAAGAGCAG ATACCTTCCAC TTATGCTTTA TCCAGAGAG 960
CTCTTGTGCT CTTCTTGTGTC ATAAAGTTGT CTCTTACCTA ACCGATTTTT ACCAATTTGA 1020
AGTAAAGGCG GACAACTAT GGAAGTGA CTTCAATCCA TTGCACTGAG CCACATTTCT 1080
CTTTTCTATA TAAAGAGCCC CATTAATATA GCTTACGGGT AGTTTGGAAC AGCTATGTTT 1140
CATATTTTCA AGATGTGAC GACCTGTCT TACTCATCAT CATTCGATCA ATCCATTTGA 1200
CTCTTTGACA AAGGGTGTAT ACTTTTCACT AAAAAAGCT ACTCTTCTG TTGATGTTTC 1260
TTTTTGTGTT TTACTTGTG CATTTTCCAC ACTATGATTT TTTTATATTT TTTAGAGGAT 1320
CAGATTTTAG CGCTGGAAAA TGAATTTCAA AATTTAGTG TAAATGCTATA AGCATTTTGG 1380
GATAGAGATA TTTTTTTTTT CTTTGGAAAC AATGTACTG GGAAGAAACA CAGATGCGCT 1440
TTGCTCTGAG TTTCAATCTG ATGATATGTA CCAATGAGCA TATCTTATG TAAAGGTTTA 1500
ATGCTGTTTA CAGTGTGATA GATAGGCGG AGATGTGAG AAGCGGGTTT TTCTCTATGC 1560
TAAATGTGTC TACTAGAGC AGCACTTCTT ACTAGTAG CAGATCTATA GCGGCACTCT 1620

5 GATGAGCTGC TACTGTGANT ACATCTTCCT GACTTAGGGA AAGCAGACGA AAAAAATATA 1680
AAGAAATATG CTATTTTCTAT ATGTGTGATA CTGAGAGAGC CATGTGATTC CTAAATATATA 1740
GGTTTCTCTT TTCTCTGTA TTCTTAGGGA ATTGCATTTA TTCACTACAT TACAAACCAT 1800
CACTGATGTA TCCAAATATG CACACATATG TCAATATGAA AATAGAGGAA TAAATATCTGT 1860
TATAGCAGAG TCAATATATG ATTCTTTCTT GTGTTATGAC ATTATCTGAC TATATATATA 1920
CCTGTATTTT TATTTATCTT CTATCAATTT TCTCTACGAA TTATGTATTT TCAATGCTCT 1980
ATAGAGTGA ATATGAAAT TATATTTCTT TTCTCTGAAA AAGAGTTGCA ACTACTTTAT 2040
TATATTTAGA AATCAATATA ACTCTTTAT ACATTTAAAA AAAAAAATA AAAAACTGAA 2100

10

15

20

25

30

35

40

45

50

55

60

(2) INFORMATION FOR SEQ ID NO: 172:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1930 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

30 CATTGATATG TGCTGCGGG TGCAATATG CAGGCGCTCC GCGCGGCTC GTGCTGTGCC 60
CGCCATGCGA CTGTGCGGG GCTTGCGGG GAGCTGGCT GAGCGGGGG CGCGGGGGGG 120
GGTGTGCTG GTGGGGGGG GCGGCAATG CTGCGAGTC CTCAGATATC TGTGTCTGAC 180
CGTTTCTTCC CACATCGACC TCAATGATCT GATATATAT GATGTAGCA ACCTCAACAG 240
ACAGTTTTTG TTTCAAAGGA AACTGTGTGG AGATCAAGG GCAAGGTTTG CCAAGGGAAG 300
TGTACTGCGA TTATACCGGA AACTAATAT GTTGTGCTAC CATGACAGCA TCATGAAACC 360
TCACTATAT GTGAAATTTT TCGACAGTT TTTACTGTT ATGATGCTTT TATATACAG 420
ACCTGCGCGA AACCATGTTA ATAGAAATG CTGCGAGCT GATGTCTCT TATATGAGG 480
TGAACAGCT GGTATCTTG CACAGTATAC TACTATGAA AAGGTGTGCA CCGAGTGTTA 540
TCACTGTAT CTTAACCGGA CCAAGAGAC GTTGTGCGC TGTACATATC GTACACACC 600
TTCAAGACCT ATACATGCA TGTGTGCGC AAGTACTTG TTCAACAGT TGTGTGCGGA 660
AAGAGATGCT GATCAAGAG TATCTCTGGA CAGAGCTGAC CTTGAGCTG CTTGCGAACC 720
AACGAGGACC GAAGCAGAG CTAGAGATC TATATGATG GTTGACATTA AACTATATTC 780
TACTAGGAA TGGCTAAT CACTGATATA TATTCAGTT AACTTTTTTA CCAAGCTTTT 840
TAAGATGAC ATCAGTATC TGTGATATC GCAAAACTA TGGCGAATA GGAACCTCC 900

5 AATTGCTGTC ATATGGAATA AGATGAGCC ATCTGCAATG GATTTTGTCA CTTCTGCTGC 960
AAGAAATGGA CCCAGTTAG GCTTGAAGA CAGCAGGTT CTAGATGTTA AGAGCTATGC 1020
AATCTCTTTT TCAAGAGCA TCGAGACTTT GAGGTTCTAT TTACAGAAA AGGGGATGG 1080
ACGTGAGCTC ATATGGAATA AGATGAGCC ATCTGCAATG GATTTTGTCA CTTCTGCTGC 1140
AACTCTGAG ATGCAATTTT TCAATATGTA TATGAGAT AGATTTGATA TCAATCAAT 1200
GGAGGAGAC ATATATCTTG CTATGCTAC TACTATGCA GTAAATGCTG GTTTGATAGT 1260
ATTGGAAGTA TTGAGATTTT TATCAGGAAA ATAGAGCAG TCGAGAGCAA TTTTITTTGA 1320
TAAAGAGCA AACCCAGAA AGAGCTTCTT TGTGCTTGT GCACTGATC CTCCAAACC 1380
CAATGTTAT GTATGTGCA CAGGCCAGA GTTGACTGT GGTGTGATG TCTATAGT 1440
GACTGTCTC ACTTACAG AGAGATAT GAAGAAAAA TTCTCTATG TAGCAGCAGA 1500
TGTCAATAT GAGATGGA AAGAGCAAT CCAATATCT TCCGAGAGG GAGAGAGGA 1560
ACTATATAT CACAGAAAT TGTCAATTT TGAATTAGA AATGCAACC GCTTCAAGC 1620
AGATGATTC CTGAGGACT ATACTTTAT TATCAATC CTTCATATG AAGACTAGG 1680
AAGAGCTTT GATTTGAG TTGTGTGTA TGTGCGGAA AAGTGGGG CCAACAGC 1740
TGAAGATCT GCAAGAGCA TAAACATG GCAATGATCA TGGAGCTTC AGCCCTCCAC 1800
CTTGAGCTC TGAAGAGC AAGATGAGG TTCTCATAG TTGATCGG TGAAGATAT 1860
TCTCAATTA TTGCGGAGG TTCATTTAG GAAGAGAG GAGGCCGCC AAGAGAGGA 1920
TTTAGATATG 1930

(2) INFORMATION FOR SEQ ID NO: 173:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1509 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

50 GCGCTGCCC TCTGTGCTGA GCTTGTCTAG GACTTCGGG TGGCTCTAAG GGGAGGAT 60
AGGCTGCGG AGCGCGGCC TTGCGCCCTG ACCAGCCCT TCTGTGCGG GTTCAACCC 120
GATCGAGTG GTACGTGCT TACGCGGGA CAGTACTG ACCGACTCT TCTCTCAGA 180
CCTCATGTC GTGCTGTCT CTCTGAAAG CAGGCTCTG CCGAGCCTG TTGACAGGA 240
CTTCTACTCC GATTTTGA AAGAGACAC AGGAGATG GAGAACTAG AACTGATCA 300
CTTATGTC GTCAAGTTTA CTAACCCAG TCGAGAGG AGTGGAGC TCACTTAC 360

5 TGTGGCCCAA AAGATGGCTG AGCCAGAGAA GGGCCCGAGCC CTCAGCATACC TGCCTTAGCT 420
GCAGGGCTTC CAGGTGGGCA TGGCAACCCC TGGGTGGCTGC AGGGGGCCCCC TGGCCCCCAA 480
GACACTCTTG CTCACCGACT CCGAGATCTT CTTCTGTGAT GAGGAGCTGTG TCCACTACCC 540
ACTGCCGAG TTTCGCAAG AGCCGCGCCA GAGAGACAGG TACCGGCTGG ACATGGGCG 600
CCGGCTCCGG GACCTGACC GATGTGCTAT GGGCTACAG ACCTACCGCC AGCCTCACC 660
CTCTCTCTTG ATGACGTCCA AGCTATGAC CTCATGGCCA GTGTCAACCT GGAACACTTT 720
GGGAGGTGC CAGGTGGCCC GCTTAGAGCC AGCCAGGGCC GTGAGGTCCA GTGGCAGTGG 780
TTTGTCCCAA GTCTGTAGAG CAGAGAGAG CTCATCTGCG TTGTGGCTCG CCAATGGGAG 840
GGCTCTGTG GCGCTGAGCT GCTGTGTGAG CTCACCGGCT AGCCAGAGCC ACACCGAGCC 900
TGTCTGTGCC AGCCTGAGCC CTTACTGGGG AGGGACAGAG GCTTTTGTGT TCTCTAAGAA 960
TGTTTTATCC TGGCTTTGCT ACCTTAATTT GACTGTCTTC GAGAGAGATG TGACATATGG 1020
TGTGTATTGT GTTAATTTCT TCTCATTTTG GAGGTGAGAA TGGCGGGGCC CTCAGGGCTG 1080
TGGGTGTGCT GTACGGCTCC CAGAGTGTGT ACAGCGTGC ACACAGATGT GGTGTCTGCT 1140
GTGTGGGAC GGTGTATAC AGCTGAGACT GTGGGTCTGA CTTTCTCTTC TACAGCTGCT 1200
TTCTGAGGT GTGAGTCCA GTCTTTGTT GCTGTGTG TGCTGTGTC TGTGTGCTTT 1260
GGCATCTTC TGTAACTCT GAGGTGTGTA GCGAATGCA CATTGGAGC TCCCAACCCA 1320
TATTGTCTT CAAGTGTGAG GTCTCCCTG ATCGAGCAA GTGGGAGAG CCGTGGGGGC 1380
AGGGAGCTG GAGCTGGCAG CACCAAGCT GATTCTGCT GCTGTATTC TCTAATGCAA 1440
TAAGCGAGG TTTCACCG TCAGAGAGAA AAAAAAAAAA AAAAAAAAAA ATTCTCGCG 1500
CCTGAGGG 1509

45 (2) INFORMATION FOR SEQ ID NO: 174:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3173 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

5 AGAGACGCG GCGATTAA ATCATTTCTT GCACTTCTAA AACATCTGTT GCTTATTTAA 240
CACCAGAGA CTCAGAAACA GCACCGCTAG TTACTGTGAC CCACAGAAAT TATTATGAC 300
GATACATGCG CTTAGGTGCG AACTACATC TGTGTATTAG GATCTGAGG ATGGAGAGCG 360
CAGGAGACTA CAAGGACAGC ATTAATACAC AGGCTGATCC CTACAGCAC ACCAGGCGCT 420
ACAGCTGCA ATCTATGCT GCGTTGGA AACCAAAAT TACACAGAT TTAAATGCAAT 480
CTGTGACAG CACTGTAAAT GTACACTCTA CATGCTCTGT AGAGAAAGAA GAAGAGATG 540
TGACATACA TTGGATTTCC CTGGAGAG AGGGTAATGT CTTTCAATC TTCCAGCTC 600
CTGAGGACA AGGCTGACT TACAGTGT TAAGCCAGAA CCGTGTGAC ACATTTCTG 660
ACTCATGTC TGGCGGGCAG CTCTGTGAG ACATGCAAT GGGCTTCTCT ACTCAGACA 720
CGGGTTGCT GAGGTGCTG GCTATGTTCT TTCTGCTTGT TGTCAATCTG TCTTAGTGT 780
TTTGTGTCG TTGTGTCAG AGAGAGAG ATCTGCTTC AAGAGAAAC ATATACACAT 840
ATATCATGCG TTCAAGGAC ACCGAGCAG CAGATCTGAG ATCTATGAT GAATCTGTC 900
AGTCCAGGT GCTTCCCTCC AAGAGAGCC CAGTCAAC ACCTTATTTCC GAAATGCAAT 960
TTGCTGATA GATGGGAAA GCGAGCAGC AGAGATGTA ACCTCTGCG ACTTCAGCT 1020
ATGAAATGCT GATCTAGGCT GCTGGCTGA ATTCTGCTC TGGAACTCA GTTACAGCA 1080
CCATTAAGCG CAGGTTCCTT GATTCAGAT CTTTCTGTC GAACTCTTAC TGGAGATG 1140
CAAACTGCA CATCTGACC TGTAGCAA GCGAGAAAC TTCTGCTGCG CATAGCTGT 1200
GCTTAATGCG ACATATGAT GATAGCTT CTTGAAATGA CTTGCTCTG ATGATATGAC 1260
AAGCAGGTT ACTTAGTATA GTTTTCCAA ACTTCTTCC ATCATACAC ATGTAGAAA 1320
TAAATTTTTT ATGGCACT GCGTAACA AGCAGATG CTCAGTCTG GAGCTGCAAT 1380
ATGACTAGG GCTCTCTG ACTGGAGTA ACAGCCCTG CCAATTAAGT TGGGAGAG 1440
GATCAATAT TTTCACACC TGTATAGG CATGCAAC CAGCGAGAT GCTCTGCTCA 1500
CAGTCAAT GTGTAGAT CCGTGTGCG TGGCTTTC CAGCGATCTT GAGCAATTA 1560
GGAAATGTA CCGTTCCTT GAGGCAATG CAGCCCTTCC CCGAGTCCA TGGCTTGGAG 1620
AGCAGATGT GGGCTGATA TAGCAACT CATTCCTTTG TCTGGAAATC TTGTGTGAG 1680
GCTAAGAG CTTAGTAGT CCAGACAG ATGACATGCT TGTGTGCTC TCTGTGAG 1740
TTGTGCTCT CAGCGATGTA GACAGCTCT CAAATGAG TTTGAGAA TTTTCTTCT 1800
GCAGGCTCA GAGATGCTG GAGACTTTT CTTGGATGCT TACTTCAAA GCTTTTAGG 1860
ATTCTTCT TGGCAGAT TTCTTCTGT ATCACTCAA GCAAGCTCAG CAGAGAGC 1920
AGCCATGCC AGTATTCGA CTCTCAAA GAGATGACC AGCTTATATT TCTCACTT 1980

CTGGGAACT GGTATATC CAACTTAA ATTAGAGAC CTTCAGAGA GCAGATCAT 2040
TCTCAGAG GAATTTGGA GATGATGTT CAGATGATCA AACTGGGTTT ATCCAGTTT 2100
5 CAAAGACTA GAGACTAGA GTTTAGGCTG AGCCAGAGTG CCGCCAGCTT GCGTGGCCC 2160
AAMAGAT CACGAGCGAG CTTCAGAGG CATTAATCTT CCTCATAGAG GAGATATAT 2220
TCAACTGTA GCGAGCATTT CATATGATCA TTTTAGGAG TGTTCCTCTT ATGTGTAGC 2280
AAGTATATC GGTAACTCC TAAATCCAA TGAATATCC TAGCCTGAGC AGCAATGGCC 2340
TGCATTAGG CAGATAGAG CATCACTGCC AATATATGAA TCCATAGACT CATTCAGAC 2400
15 CAACTAACAT TAGCAATATG TTAGAGAGCT GAGGCCCCA AAGTAGAGAA TGTGCAATAT 2460
GTCTGCTCTT GTTAGCTCA GAGACATTT CAGAGACGA CACTACATTT AACCTGAC 2520
TCAAGTCCA AATATATGCA TGAATGCTA GAAATATACC TTATGAGGGG GCGAGGCTCA 2580
AGCTGGGCTT TGAAGTGG ATCAATATTTG GATAGAGAT GAGCAGAGCA GAGGCGCTCC 2640
AATGAGAGA AGCTGAGAAA ATGAGAGGG GCTTGATCA GTGGGGTGTG TTGAGAGCAC 2700
25 GTCTGCAAT GCAATGCA TGGTCAATGT CCGTTGCTTA TGTGTGGCAG AGTGTCCAG 2760
CGAATATGT GCGCCGAGCC CATGTGCAAT TAGATGTCTT TCAATGCTCA CTTGAMAGG 2820
TACTCTTCT GTTAGGCTTT CCGTGTATC AGAATCATA ATTATATAGG GATCTTTTCA 2880
CACTGCTGTT TTTTCTCTTT TGTCTCTTCT ATCACTAATA CTCATCTCAT TCAAGCTTAC 2940
AGCAATACTA ATTAATTTT TTTCTCACTA CATGTATCAT GTGGAAATTA CAGATTAAGG 3000
35 GAGAGCTGCT GCGGTGTGTG GTCAAGCTTG TAACTGCAAC ACTTTGGAG GCGAAGGAG 3060
GCGATACCC TGAAGTCAAG AATTGAGAT TATCTGTGCC AACATGTTGA AACCCCATTT 3120
NFACTAAAA TACGAATTTA GCGAGGTGTG GTGCGACGA TCTGTATGTC CAG 3173

(2) INFORMATION FOR SEQ ID NO: 175:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 991 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

AAATTTGGA CAGCTGAG GAGAGAGAG GAGAGGCGG CAGGACCAA GTGAGAGCA 60
TGAATTTACA GTGTGTTTC GTTGTGCTCC TGGGTACAT ACTGATATTT TGTCTAGTAG 120
AAGACAGG TCTGAGAGA TGTGTGATTT CAGAGCAT GCAATATATA GAGAGATTT 180
TCCAGAAAT CAAAGAGCC ATTCAGACTA AGCAGACCTT CCGAATGTC ACTATCTGTT 240

CCAAATTTGA GACTCTGAG ATCAATTAGC CTTAGATGTT GTGCTGCGTG ACCAAGAAC 300
TCTTGCGTTT CTAGCTGAGC AAGGTGTTCA AGATGATCA GAGCGAAGC CCGAATATCT 360
5 TGAAGAAAT CAGCAGCAAT CCGAACTGTT TCTCTACAT CAGAAAATCT CTGCGGCAT 420
GTGAGACA GAGCGATGTT CACTGAGGC AGGAGGCCAC GATTCGACC AGATGATATCC 480
10 ATGCAACTA TGAATAGCTG GAGTGCAGC CTCTGCCAT TAAATGCGTG GAGAGAGCTG 540
AGTCTTTCTT AGCTTGATTT AATGAGATC ATGAGATAT GTCTCTAGCT TGAATGACAG 600
GAACCTGAT ATGATGCGC GATGAGAC CCGCTGTGCG GTTACTGTT GAGAGAGCC 660
15 CACTTGAG GCGAAGGAGA TGGGAGGCC CCGTTGAGC TGAAGATGCC ACTGCTGCG 720
CTCAGGCTGT CTATTTGCGC TTGAATATAG CCAAAAAGTC TACTGTGTA TTTGTATATA 780
20 ACTCATCTG CTGAAGGCC CTGAGGCCA TCTTGGAAT AAAGGGCTGCT CTTCGCAATCT 840
AATTTATCT GAGTCAATAT ATTCATGTC TGTGATGCA GCGAAGTAT ATCTGTATCT 900
25 ACACATTTTA CTGAGTGTCT TTTCTGATTA AATTCATAT TTTACTTTAA AAAAAAATA 960
AAAACTGA GGGGGGCC GTACCAATTT T 991

(2) INFORMATION FOR SEQ ID NO: 176:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1290 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

ACAGGCTCT TCGAGGCTG AGCCGCGCTC TCTCACTCA CTTCAAGCCC CAGGGGCGCC 60
CTCAGAGGG CCGCTCTCTT CCGTGCAGCG CTCTGCTGTT CTGCGCTGCC CTTGAGAGAG 120
45 AACAGGCCA TGGTTCGCCC CCGTCTGCTG CCGCTGTGCG YCGTGTGTCM GCGCGCAGA 180
TTTTCGAC CTGTGTGCTC CACAGATCT GGTCAAGCTT ACCTTATGCG GGTCACTCAA 240
CGAAGACC TTCAAGCTC CATGGTGGC TCTGTGGAAA TCCCTTTCTC CTTCATTTAC 300
50 CCGTGGAT TAGCAVAGY TCCGACCTG AGAATATCT GAGAGCGGG CCACTTCCAC 360
GGGAGTCTT TCTAGACAC AAGGCGGCTT TCAATTCACA AGAATATAT GAGCGGCTC 420
55 TTTTCAAT GAGAGAGGG TCAAGAGAGC GGTCTCTCA GATCTTCAA CTTGCGGAG 480
GAGAGCAAT CTGTATTTT CTGCGAGTC GAGCTGACA CCGGAGATC AGGAGAGAG 540
60 CAGTTGAT CCATAGGG GACCAATCT ACATACACC AGGCTGTGAC AACCAACAC 600

5 ACCTGAGGC CAGGACGAC AACCAACA GTGGGCTCTA GGTTCACAGA AACCAAGGG 660
CACTCAGAT CATGGCACT AAGTCTGAC ACTGCCATCA GGTTCGATTT GGTCTGTGCT 720
GTCTCAGAA CTGTCTATTT GGGAGTGTG TGCTCTCTCC TCTGTGTGTG AGGAGAGGA 780
AAGTACAG GGGCCAGGC AGTGACTTCT GACCAACAGA GTGTGGGGAG AAGGATATG 840
TATTAGCCC GAGGAGATG ATGTGAGAC GGTTCGTGAG TCTTCACAG TGTTCCTCCA 900
TTGGCAGAT ACATGAGAG CACCTGAGG ACCTTTAATA GGCNAAGCG CAGGACAGAA 960
GGAGCTGGG TCCCTGATC ACCGACTGGA GAGAGTTAC CTACAGAGC CTTCATCCAG 1020
GAGATCCAC ACTGCATCA TATAGGATG AGTCTGAGC TCCACTGAT TAAACCACTG 1080
GCTTTGGGG GCTGTTATTT ATAGAGTGC AAGAGTTCC TTATCTCTCC CAGAGATGG 1140
AAATACAGT TTAATTTGCT TACCATGAC CCGTTATCTC GTGTCCACA TTTTCCATTC 1200
TGTATGTGCT CTGTCTCTA TCCGAGAGG TTTTGGGAA TAAATAGCT GAAATGATTC 1260
TCACTAATA AAAAAAAAAA AAAAACTGCA 1290

(2) INFORMATION FOR SEQ ID NO: 177:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2290 base pairs
(B) TYPE: nucleic acid
(C) STRAIN/GENES: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

5 TGGGGCCCTT TTGGATGCT CTGGGTGTTT TTGCGAGAG TTACAGGATG TCAGTGTGG 60
GGAGCTCAG ACCCTTGCTG TCGACCAATG AAGGCTGTTC CAGACCAAGT GCTTCAGAC 120
ATTTCAGGC TCGAGGAG AGGCTGGGAG CCCCCACAGA AAGCACAGA AAATGCAGAA 180
AAAAACAT CTTTTATTTT TTTTCTCTT TTATATGAA ACAAACAAA ATGCCCCAGG 240
AGAGGGTCC ATGATACCA GAACATCCA AAGATACTTT CTACCATTT TATCTGTGTG 300
TGTGAGGCC AGCATGCAA TAAACAGCT AAATCTACTA CATTGAGCT ATTTCAGTA 360
ACTGACATTT ACAGATAT ACTAGAACG GCATTAATA GTTTAGAAA AGTACAGGTA 420
ACTTGCATG CACATCATC AGAAGATTA CATTTAAT ATAAAAAGA AAACTTCT 480
GGAGCATTA TCCAGTATT AAGGACAGT GCTACTCTGG ATGTGACAAA TTCTGTATGT 540
GGGTGTACT CTTTCCAAA AGCATCTGAG AGGCTGAGT GCTGCNAAG AACACACCA 600
AAACACACA CACAAAAA TGTGTCTTAC AGTTGTATAG CAGATGACA CTGCCACACA 660
CAAGAGGAGG TCTGGATTC AGTTCAGCC GAGGCTGC CCGCTGGCC TCCAGGGTTC 720

5 ATTGAGATG TTCTCAATC CAAATTCGAC AAGCAATG TCACACTGCT TCCTCCCTTG 780
AAAAAGCAT GTTAGAGCT GGTTCACAGG TCTCAGAGT GGAACAACT AATTGAATCA 840
CCGAGCTTT CTAAACAGA AAAAAAGAC GTGACTGCA CCGTCAGCC GCGAGCAGG 900
GGCTGAGCA AATCATTAAT CTTCTGAAC TCTGAAGA AACAGTGT GAGTCTGGA 960
CAAAAACTT TAAAAAGT ACAGACCAA TCGAGTGT CAGTGTAC CCCTGGCT 1020
GTCAAGGTTA GTGGCTCTT TCTGAATCAA AAGAGCAGAG GCTAGGCTAC GCGACCTCA 1080
CAGAGGCA GCGCAGAGG AAGGCCCA CAACTGCTCC TGTGGCTG CCGCAGGCC 1140
GGGGCTGCC CCGAGGCTC ACAAAGGCTC CACTCACTG CCGTGTGCT GAGGGGCTAT 1200
TCCAGAGAG GCGCCGCGC CCGAGGACC CCGGCTTAG GGTGTACGTA TCAGCGGCC 1260
CTGTCTGCG AGCAGGTTAC CAAACAGCT GCGTGAGAC CTGTCACTG TCTGTGTGA 1320
ATTCTTAA TTGGTTTAA ATAGTCAAT AAGATCTGT TTAGAATA CTTTGAATA 1380
CGAGGTAC TTTAAAAAT GGAACCTTC AATCCATTT ATATTTTAT TATTAACAAA 1440
ACTTAATTA AGTTTACA AACTGCTCA AACTCACA AGTTCAGAC TCAGCAGAA 1500
TTAAAAAAT GATATTTAC CAGATCTCC TCTCAGAGT TCCTCTCCA GTAAAGGTAT 1560
ACCTACATC GTAGGCTCA GTGACTCTG AATCAATTT ATGTGTGTT TAAATATCAG 1620
GTGTATGAG ATACTATGA TAGTCTCAT ATCCATCCAG AAATCTGCG AAAAAAGCT 1680
GCGCACTTA CAGACAGAC CACACACAG CTCCATACC AGGCTGCG TCGAGGCTGC 1740
CCAGGCTGA GTTCCGGAG AATGCTGCT TTAGTCAAT TCCTGACTTA CTGTACAC 1800
GCTTCAGAG GAGCAGCTT GCGAGTCTC GGAAGGAT CCGCTGCGG CCGCAGACT 1860
CTTCACCCC TCGGAGAGGC AGACAGGCTC GGAAGGCT GCGCAGGCA CTGAGGCTG 1920
CGAGGAGCA GCAATGTCCA CCGCAGGCC TCGAGGCTA ACTGTGCA TCGTGGCGG 1980
AGCGGCTAT CTGATGTG GCGAGTGTG GGAATGCGG GTACTCAG ACCATATGCG 2040
AGCTCTGGG AAAAAAAG CAAACATCA CTTCTATTA AACTGTGAT ATATATATT 2100
TTTACATAG AAGTTAAA ATCAGACTT AGATTTACTA TACATTTTTT CTCTCAGATT 2160
ACAAAGTTA TATATATA CTGGGTCC CTAAATGAT TTCTTTTAAA AAGTCTTAA 2220
AAGAGCAGA AATGATACA AAGACATA ACAAATAAA AATTAGAT GTCTGTAGC 2280
TGAAGCTGT 2290

(2) INFORMATION FOR SEQ ID NO: 178:

60

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 549 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

5 GCGACGAGCC ATGCGCTGGC TGTCTGTGAT TCTTACAGTC ACTTTGTTGG CTGTGTTGTA 60
CTGACGAGCT ACTGCAATG TGGCAGAGG ATGACTTATT CTTTCTGAG AGGCGAAGAA 120
TGTGGAATAG TGTCTGTCCA TGGCTCTGCT CATGGCTAC CAGCTCTGGC AGCTGTGTTA 180
15 ATGATATACA ACGAGAGAG AGCTGCTGAG AACTGACTTC TGGGAACTCC CTGGGAATGT 240
TTGTGTCAGG AATGTATGAG GATATAGCT GATTGCTGTG ATCTGGGCCC TGTGTGATGT 300
ACTGAGAGGG TAAAGGGCCA CATCTCTGTT GACTCTGGG GAACTCATTC ACAAAGAGAA 360
TATTTCGAG ATGCTTCTGA AGATTGGCTA AAATATAGCG GTTTTCACCC CCGTGAATCC 420
ATCATCTTA GATGCTGCT TGCAGGAG CAGGAAGCTG ATTTACAGAA GTGACATGAA 480
25 AACATTCAT CCGAGATTT GATGAGTTC AATTATGTT TCTAGCTATT AAAGAGAAA 540
AAAAAAAAA 549

(2) INFORMATION FOR SEQ ID NO: 179:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1509 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

40 GCGACGAGGG CTCATTGATT CCGCGCGGG CCTGCGAGAC ACTTGCGCCC TTCTGAGACC 60
GCGCGCGGCA TCGCGCGCGG CAGCGCGCGG CATGTGCGGC CAGAGCTGCG AGACGCGCTC 120
CGCATTCGAG ATCTGCGCGA TCACTGCGGC AGATGATGCC AACGTGCGCG GCAATGTCCA 180
CGGGGGGAGC ATCTGAGAA TGAATGAGAA GCGAGCGGCC ATCATGAGCA CCGCGCATTT 240
50 CAGACGCCAG AACGGGAGC GCTGTGTGCG CGCTGTGCGT CTTGTGAGC GACACGAGTT 300
CCTGTCTCCC ATGTGATGCG GTGAGTGTGC GATGTGTGAG GCGAGATCA CTTACAGCTC 360
CAGGACTCT GTGAGGTGCG AGTCTAGCT GATGTGCGAA AAGTCTCTCA CAGGTGCGAA 420
55 AAAGCTGACC AATTAAGGCA CCGTGTGTTA TGTGCGCGTG TGTGCGCGTG TGTGCGCGAA 480
GCTCTCTGAG GTGCTCTCTG TTGTGTATTC CCGGAGAGAG CAGGAGAGAG AGGCGCGGAA 540
60 GCGGTATGAA GCTCGAGAG TGAAGCGCAT GCGAGACGAG TCGAGGAGCG GCGACATCTG 600

5 CAGCCAGTC CTCAGCCAG AGCGAGACAC TGTCACTAC AGCGATCCA GCTTGATCCA 660
CTGTGTGGG CTTTCAAGCT GCACTCTGCA CGCTTTTGTG CAGCGAGGTG TGACCATGAA 720
GCTCATGAT GAGGTGCGCG GATCTGTGCG TGCAGCGCAC TGCAGACCA ACATCTGTAC 780
AGCTTCTGTG GAGCGCTTTA ATTTTCTGTA CAGATGAGA AAGGCTGCG TCAATCACT 840
10 CTGCGAGCC ATGACTTTCA CGAGCAATTA GTCCATGAG ATCCAGGTGT TGGTGAAGCC 900
GAGCGCTGTT GTGACAGCT CTGAGAGCG CTAGCGGGCC GCGATGTGCT TCTTCACTTA 960
CTGTGTCTG AGCCAGGAG GCACTGTGCT GCTGTGTCCC CAGCTGTGTC CCGAGACCGA 1020
15 GCGAGAGAG AAGCGCTTTG AGAAGGCGAA AGGCGGTGAC CTGAGATGA AGGCGAAGCR 1080
ACAGGGCAC GCGAGCTTC AGCCGTAGAC TCCCTCTCTC TGCAGCTGCT GCTTCAAGTA 1140
20 GCAATGCGAA CGGGCGGAGT GTCCATCTAC TTAGAGGTTT CCGCTTGTGC CAAMACCCA 1200
ATTCAATTC AGAGCTGTG TTGTCTGAG TTGTCTGATC ACAGTGTTTA CTTGTACTCT 1260
CTCTGCGAA CTTACAGCC AAGCTTTAT TTATATCAT TCAATATCAA TCTTACAGAG 1320
25 TTTTGTCCC AGCGCGGGA GCGTTGGGC AGAAGCCCTC GCGAATGCTT CCGAGACGC 1380
TGTAGGCTAT GCGAGAGCC CAGCAGCTCT AATTAAGCTG CTGCTTGTGCT GGAAMAAAA 1440
30 AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1500
AAAAAAAAA 1509

(2) INFORMATION FOR SEQ ID NO: 180:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1316 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

AGCTATATCA TAGGAAGAT GCGCAGACG CGGTTACGAG TAAATGTGTC TCGGGCGAG 60
CGAACCCAG TCCGGGGGGC GCGCTCAGCC TGAATTTCCAG CCGCAAGCC AGCAGCGGCT 120
GCGCTCCGG TTCTCGGTGC GCGTCCAGCC TGCATCTCTCA GACCTTCCGG CAGCAGCGGC 180
TGAAGTGGG GCTCTGTGCC AGACCCCGGC CTCAGCGCAA ATCAGAGCCA GACCCGAGCG 240
55 CCGCTCTGC CTGGTCTGCG TCTTTCAGGG CCGTTCGCGG CGGGCGCGCT GATCAGGCTG 300
CAGCGATCA TTGTGGCTC CATTTGGAC AGCTAGAGAA CAGCGACCA GCGTGTCTCC 360
CGAGTATCG GCACTGTTT GCGACTGTC GACAACTCT CAGTGGAGGT CACCAATTCG 420

TTTTTGGTC GGCACATCA CTCAGAGAT GAGTGGCTG TTGACATGA ATTTCCTAG 480
AAATGTAAG AACTGATTA AAAGATTTCT CCAATGAGC TCATCTCTGGC CTGTACGCT 540
ACGGGCATG ACATGACGA GACTCTGTC CTGATTCAT GAGTACTACA CCGGAGGCC 600
CCCAACCCC ATCAGCTCA CTGTGACAC AGTCTCTCAG AACGGGGCCA TGACATGCA 660
AGCTTACGTC AGCATTTTA TTGGGTGCC TTGGAGGACC ATGGAGTGA TTTTCAAGCC 720
TCTGACGTC AATACGCTT ACTAGACAC TGAAGCATC GCACTTGACC TCATCATGA 780
GACTGCTTT AGCCCAACA GAGTATGTC ACTCTCAGT GACTTGCAGC AGTTAGAGG 840
GGCATCAGCT CCGATCCAGG ATGCCCTGAG TACATGTTTG CATATGCGAG AGCATGACT 900
GTCTGGAAG GTCTGAGTC AGATGACTT GGGGCCCTTC CTGATGAGCC TGGTTACCA 960
AGTACCGAAA ATAGTCTCCG ATGACTTTGA GACCATGCTC AACAGCAACA TCATGAGCT 1020
TTTGTATGTC ACTTACCTG CCACTCTCAC ACATGTCAG ATTGCATCA ATGAAAACT 1080
TGTAACTCT TGAATGACC CCAAGATTA CACTTCTGTC TCTAGTATTT AACCCAGCA 1140
CTCAGAGTG AAGGAGAAAT GGGTTTTTG TGGTCTTGA TCACTACTAG ATAGTCAAT 1200
GTGTGTGACT CTATTAAGG GAGCTTACT TTTGTAAATT AAAAAAAMA AAAAAACCN 1260
SGGGGGGGG CCGGTCCCA TTSSGCTTT KGTATTCGT NTTAAATCC CAGGCC 1316

(2) INFORMATION FOR SEQ ID NO: 181:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 777 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

GCATGKCA GACATGACTT CTATGGCAG GCTGTGACG TGGAGGGTC ATGAGGCGA 60
CATGATTAAG GGTGCTCTTT ATGCTGCTTG CTCTGGATC CACAGAGGG CTATCTGCT 120
TTATGGGCTT GGGGACTAGA ATTGATGCT TCAAAACAT CACTGTGTGG CCAACAGTT 180
TGACCAAAG GTAGATGATA ATGCTTTCA GTGCTTAGA GAATACTAC GTTATTAAGG 240
CAATCTATT GGGACTGAA CTTTGAAGC CACATATTC AAGAGGCTT GCTTACCTGT 300
TGGGGGCCAA GAGGATGTT ACCAAATG GTTGAGGAA YTTGGTGGG AMCARRGGG 360
GKGGGARRH CHRGGGTTG SCATTTCSK KGGWACCTT TTAGGTTAAR RRGGGCKGTM 420
ATTAGATTT GGTAAAGTA GATCTTTTG CCGTGCAMA TTTGCTGCTT GGTGAAATGY 480
TCTTGTGTC CTATACCC CTATACCTAG TATGTCCTCC ACTTACTTTC TCATTAATG 540

AGATAGAAA CTCCTGTGAT AGGAGAGTG AAGAGGGAT ATGTGCTAGA GCACTTATTT 600
TCAGTGAAT GCTCTGCTGT AGCTTTTCCA TTCTGTGAG CTGGCTTCC TAAATATTC 660
AGGTTTGTA GCTTGAGGA GAACTTTGAT GGAAGAGAA CTTTCCCTTC TGTACTGTTA 720
ACTTAAAT AATAGCTCC TCAITCAAG TAAAAAAA AAAAAAAA AAAAAA 777

(2) INFORMATION FOR SEQ ID NO: 182:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 791 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

GGCAGATA ACTATGACA TGTATGCTT AAATATTTT TTAATTTTA TATCTGCGC 60
ACTGCTCTC AATGCTGAC ATGTGTGCA GGGAGCAAT GCTTCTTGT TTCTGAATTT 120
GZTCTTTTAT ACTGCTTTT TTTCCCATCT TCTCAGCTCC TGCCCTCTCT TCAGGTTACT 180
TCCGTGCCA GAACCTGCC AGGTGAGAG CAGAGAGAA GCTCTATGG TCACAGCAC 240
AGATGCGC TGAAGATTA TTAATTTGT TTTGGCTTGA ATTTTCTGTA TGGTTTACT 300
GATCTGGA AGAATATC TTGCGAGAA AATGATAGN CTTGACAT OTTGATGAT 360
CCTGACAC CTTGAAGAC ATTCTTAATA TGGTTTGTCA GGCAGATG TTAATGATCA 420
TTTGTGCTT GAGGTAGAG TCTTGAGAA TCACAGACT TCATGATTA AATGCTGACT 480
TGCCCTTGA CTGGCTCTG TGAGATGGC CTCTGTGACT GTTACAGTA GZTGTGACA 540
CACACACT ACAGGACCA COTGTTGGC TGTGACTAG GGGCAGCT CCGTCAAGC 600
CACTATTA TATCAGCTT TTACATGGG CTGTTTCA TAATCTGAT GAATCTGAT 660
TGGTTCTTT CTTGATACC CTTCTGCTT TTCAATGTTT TTTGTTAGTT TTCTGCTTT 720
TTAATTTTA CTGAAATA AATCTTTAG GCTATTCACC TTAGCTTAT AAAAAAAA 780
AAAAAAAACT C 791

(2) INFORMATION FOR SEQ ID NO: 183:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

5 AATTGATTA ACGCTTGA AGAAGCTCT GGTTTGAG GCTTAGATAG CAGCACTGCC 60
AGTAGCATGG AGCTGGAGGA ACTTGGCAT GAGAAAGAGA TCGAGAGGA GGAATATAG 120
AAGCTATGG GCGAGATACA TGAAGTGA TCCGAATTAC AGGATATGCA GCGACAGCA 180
GTTAATGAG CAGATCAGC AAGAGAGAG TTACAGAGTC TCGATGACCA ATAGAGTGG 240
CAGAAAGCAT CAAACAGGA ACTAGAGCA GAACTGGAGC GACTGAAGCA GGAATTCCAC 300
TATATAGAG AAGATCTTGA TCGAGAG AGCAATATTC AAGCAGAT TAAAGATGCA 360
GACGAGAAA TTGAAACT CAGGATCAG CTTACCAATA AACTTTAG CAATACAGT 420
CAGCTGAGT TAGAAATGG ACTGATCAG CTAAAGAGA CTCTCATCA GAAACAGAC 480
20 ATCTGAGGA GTCTCAGAC AGAAAGAC TCCCTGGTCT TTCAACTGCA GCGGCTGGA 540
CAGCAGTCA ACTCCGCTC TCGAGTAGT AGTATGGGT CTTCGATTA TATCTGGA 600
ATTGACATG GTGAGGAC TGTCTGCA AATGTTCCTG TTCTTTTAA TGAACAGGA 660
ACTAATCTG CAGGATGTA GCGAAAGTT GCGAAAGCTG CTATGCAAT TGTATGTTT 720
AGTATGCTC TGGGATTTT TCTCGAAGA TACCCCAAG GCGAGTTT TGTATATTA 780
30 TATATGCTT TCTTCACT CTGGTCATG AATGTCTGT TGACTTAC ACCGAAATG 840
CAGCAGACC AACATATG CAATGAGC AAGCCAGTT GTTGAGTGA TTGTGTCTT 900
TTTTCTAGC TTGGATCTG CAGAGGCT AATGCTTAA AATTTCTG AACAGTGCAC 960
AAGATATTT TATCACTACA AGCTTTTAC TTTTATAGT ATTGACAG TATCTTACT 1020
AATCTTCCA ATTTGCTTGA AATGTATGA GTTTCTTAAA CAGCAATTA TTTAAGAC 1080
40 TCACTCTGC TTTATCTGAG TTATGTGCT CTATATATTA TGTAGAGAA GATGTGCGG 1140
TTGTCACT CTGTACAGAC CATCTATG TTAGGTGACA TTGATATGG GTTATATCA 1200
GCGAACTAA TTGTATTTAG TCGAATAAT AAAAAATTT TTTTATTA TTCACTCTGC 1260
45 TTTTGAAT TCAATATTT AACTTTGCA AAGATTTTAC TTTGTACATG TTACAGGCTT 1320
GATGTGTA AATCTTTTA TAAATATTA AATTAAGTA AATATGCA TTTTCTTTTC 1380
50 TAAAAAAA AAAAAAAA CTGA 1405

(2) INFORMATION FOR SEQ ID NO: 184:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1596 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

5 GTATCATGT GCGCGGAGA ACTGTCTCT TTGAGGCGA CCGTAGGGG CCGAAGGGA 60
AACTGGAG GGAAGGTGAC GCGGAGCGA GCAATTGAGA TCTGCTGGT AGACTGGTG 120
10 CACACAGC ATGTGGCTG CAGGCTGGT GTGTCTCGG ACACTACTT CTAGGGTTTT 180
CCACCACT TTACCAAG GCTCCGCTGT TGTGAAGAT TCCATCACA AGAATCAATG 240
GCTTTACA CTTAGCAGG AATATGCGAC CAACACAGA ATTGGATTC GCGTGGGAG 300
15 AACTGCGA GAATCAAG AGGCAATTT GGAACATCG ATCGAAAAA TATTTAAT 360
TATCAGAG GGAAGATGT TTGTGCTGG AGGCTGCTT GTTGTCTTG GAGCATTTG 420
20 CTACTATGC TTGGACTGT CTATGAGAT TCGAGTATT GAAAGGCTG TAAATTTGCC 480
TGTATATC AAGATGAA TTCAATTCAC CTATATATC TTACAGGGA GTATTTGGTT 540
AAGAGTTG TCTCCATAG CAATGAGAG AAGGCTGTT CTCAATGACT TCATGATGAG 600
25 AAGCTCTGG GTGACATG GTGTGAGCTT TCGACCATG GTTGGAGCTG GAATGCTGT 660
ACGATCAT CCAATATGACC AAGAGCCAG GCGAAAGCAT CTTGTCTGT TCGTACATTC 720
TGTGTATG GTGCTAGTG TGGCTCTCT GACATATTA GCGGCTCTC TTCTCATGAG 780
30 AAGCTCATG TACAGAGCTG GCAATGTGG AGGCTCTCC ACTGTGCCA TGTGTGGCC 840
CAGTGAAAG TTCTGACA TGGTTCACC CTTGGAGTG GCGTGGCTC TGTCTTTGT 900
35 GTCTCATG GATCTATGT TTCTTCACC TACCACTGG CTGTGTCCA CTCTTTACTC 960
AGTGGCAAT TACGTTGAT TACTTTCTTT CAGCATGTC GTTGTATG ATACCGAAA 1020
AGTATATG CTTGAGAG TATACCAAT GTATGAGTT CAATAATAT ATCCCATTA 1080
40 CTGATGCTG ACTATCACA TGTATGAT TAAATATTT ATCGAGTTG CAATATGCT 1140
GCGAATGGA GCGAAGGAA AAGATGAG TCACTAGCT TCTGGCTTCT CTGCTACATC 1200
45 AATATCTG TTTATGCGG CAGTATGCA TTAATATGT TGTACAGCA GCTTTGCTTG 1260
AAGTTTGA GATTAAGAAC ATGTATCAT ATTTAATAT TCCGATATG TGAATGCTCA 1320
GCTCTGCTT TTTTCTGGA GATTAATGC AGTATATCT TCCAAATTA GCGACAGAT 1380
50 TTTCAATCT CATGTTCAG TGAATTTAAA ATGTTTGTGT GAATGTGAA ACTAAGTTT 1440
GTGTCTGAG AATGTATATC TTTTCTTAC TTTTAAATTT AGTAGGTTCA CTGAGTACT 1500
55 AAAAAATG AAGCTGTGT TTGCAATTT TTGCGAGTG CAGATATAT TAAATATAGC 1560
ATTCATGTA AAGTATTTT AAGTTGATC TATATN 1596

60

(2) INFORMATION FOR SEQ ID NO: 185:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2293 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

5 GCGCAGGCG CGTAGAGCA GCGAGAGCAG GACAGAGGCG ACTCGAGCG AACCGCGCG 60
ACCGTAGCG AGGAGTCTCT CTCTCGGCG CTGAGGAGCG GCGAGGTTA CACATCTTT 120
TGGATGAGCT GTATATCATC TCGATTACCT GCGTGTATGT TAATTGATTT AGCATTTGAG 180
CGAGCTTAC AACTAGGCG TTGTGGATG AGACTTAAA GTCTATTAC AAGAAAGGTA 240
ATCTGTAGCT CAGATGCTCC AACAGGCGAT GTCTCTCTG ATGAGCTCT GAGCATGTT 300
AAGAAAGCTC ACCCTCGAG AACGTCAG AACTGATG ATTAATCTAG TGGTAGACA 360
TGGATCTCAT TAAATTTGCA TTATCAAGTTA AGAATGTTAC GCGAGGATTT AGCTTAAAGC 420
CTGTGGAAA AGGCTTATTT GACAGAGAG AACAGAGCT TCTACTTTT TGCATGACA 480
ACATCTCCC TCACCATTTA CACATTTAG CAGGCTCTA TCAGAGAT AGAGAGAGCC 540
GTCTCTGACA AATGGTGA TGCCTCTAC GCGCTGACA GCGCTTCT GCGCTCTATT 600
TACTGTGCTC ATGCTGCGA GTCCTGCGG AATGCTTTTG CTCCTCTCT GCGAGAGCG 660
TATGATTTGG CTACGAGAG ATGCGGGAG CTCTCTGACT TAGAGCTGTA AGTGAGATTT 720
CTGAGGCTCA ACACCAATCA GCTTCTGTGG GCGGTGTGTG GCGCTTCTAC CACTTAATCT 780
TCTCGGGGTT GAGCATTTCT CTTTCTCTC AACTTAAACA GTATTTTTTC TTCTGTGTAC 840
TTCTGTGTTT CTGTAAATTT TACTTTTCCA CACTTAATTT GCTTCTGTT TTACAAATG 900
GTGGTGGCTT TTTTCTTTTT TGTACGTGTA CAGATTTTG CTGTAGAG AGGCTTCTCT 960
CTTCTCTGTT TTAAAAAAG TTTTTACTGC ATATTGGCAT TCCATCTCT GTTCCATCC 1020
TCACTGTATC CTGTTTGGG TTCTGTGCTT ACTTTGACTT TCAGATGAC TCGAGCTTCC 1080
TCATAGGAC ACCTTTTGA TGACTTGC TTGAGTTTCT CCATATGTGC ATGTACATCT 1140
ACATCTGCG CTACGTTCA GACAGAGCTC ACAGAAAGCG CTTCATCTCA CCAAGGTAAA 1200
ATATCTGTAT CTATTAGCAG ATTTTTTACA TAGACTTTCG TTGAGATGTA TACTTTAGCA 1260
AATTAATTTT AATGTAAAC AGCAGAGTAA ATACTTAATA TAAATTTTCC CTGTGATTTT 1320
GCTTCCGAT TAAATCTATT GTATTATTAC ACTTGTATA ATTTTACTA TAAGGTGCA 1380
ATGTGTTAC AGAGCAGTT TGGATGCGC TGCATTCAT TTATGTTGTA TATAGTTTGA 1440
ATTATATATA AATTACCGTT TCTTCTGCCC ACCGCTGCTC CCACTTACT ATTTTGCAG 1500

ATCATATCAG TTGTACACT GGTGCGCTC GCTTCTCTCA ATCATGTTTA TTTCATGCGA 1560
AAATGCACT CTGTGCGTG AGCGAGAG AGAAGATGTG TGTCTATTTG GTTCTGGGAT 1620
TTTTTGAGCT GTGCCATTTA TGTACTCTT TGCCTATGCA TCCCTTTTAT AGATTTTTTT 1680
TAAATTTTAT CTTACTGTTT TTATATTTTC TATTTGGAG AGGCTTGTGA CCGATTAACA 1740
TCTTGAGTTT CTTTTCTCT CACAGTAA ATTAATATCT GCTCTGAAT GTCATTTATC 1800
TACTTACGCA TTCTTGGGA AAAAATCAA ATGTGAGTCC TAGCATATTT TGCATGTAAA 1860
TTGTAGCAA GTATATGTTA CAGCCAGAG GATTAGAT TTGTAAAG AGAGCTTAT 1920
GTTTTTATTT TTATATACA ATTAGATTA TTAGATTTT CAGACTTAA ACCTTTGCTT 1980
TTTAAAGTTT AATTTTACTA TTCTTTATC ACTTATTT ATCATGCA TTGTTTCTAT 2040
AATGTAAATA CTATATGTT AGCAATTTA ATGTCAAAAT TTTTATTTAC CATATGCTAT 2100
GTTTATAGTG GCGCTTTCAG GTGTTTAGAG ATTTTTTTTG TTGTGTTTAA CATTCTATCC 2160
AAAGTACTA GATGCTGAT AACTCTAGAG TTGAATTTTA AGGATTTCC TTATATGTTAT 2220
ACTATCTTTT TATCTGACT ATAAATAAA CATGATCTT GAAATGCTCT TAAATTAATA 2280
AAAAAAAA AAA 2293

(2) INFORMATION FOR SEQ ID NO: 186:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

GCGAGAGCG GAGCGGCGC ACCGTACCT GCGAGGTGTG GTTCTACTC GTGCGCTTCC 60
CGTGGGTTT GCGAGTTTA GCGACTATTG GCGCTGCGCC AGCGCGGCTG CAGAGCTGG 120
GCGGTGCTG CTGTGCGCG GTGATCTAG GCGGTCTCTT GCGCTTCTAG CTGTTCGCG 180
GTGTAGGTGG GAGTACGGA CGTTTCGCG CCCGAGGTGT CCGCGAGGT GCGCGCATG 240
GCGCGCAGCG GAGAGCATGG CTGAGCGAT GTTCTGGGTG GACTTGGAGA TGACAGATTT 300
GAGATTGAG AAGGACAGA TTATTGAGT GCGCTGTCTG ATAACTGACT CTGATCTCA 360
CAATTTGCTT GAGGTCTTA ACTGTATTAT AAAGAACCA GATGTGTTCC TGGACAGCAT 420
GTGAGATTGG TGTAAAGAG ATACGCGAA GTCTGCGCTT ACCAAGCGAG TGAAGAGAG 480
TACATTTACA TTGCGAGCG CAGATATGA ATTTCTGTCC TTGTAGAGAC AGCAGACTCC 540

5 TCCAGGCTC TGTCACTTG CAGGAATTC AGTTCATGAA GATTAAGAT TTCTTGACA 600
ATACATGCC CAGTTCATCA ACATCTTCA TTATGATTA ATGATGTGA GATCTGTAA 660
AGAACTATC AGAAGCTGAT ATCCAGAGA ATATGATTT GCACCAAGA AGCCTGCTC 720
TCTATAGGCA CTTCATGCA TTATGAGAG CATCAAGAG CTTCAGTTT ACCGAATTA 780
CATCTTCAAG AAAAATATAG ATGAAGAA GAGGAATTT ATGAATATG GCGAATATCA 840
GAGAACCTG AGTTATGCC ATTTATCAT CTCCACTTAC ATCTTATCT GAGGCAACT 900
TCTGTGATTT TTTTATCTC AGCTGATG CTTCGAGAG CACTTGTGAT TAATCTGAT 960
CTCCAGATG ATTTACTGAG CAGACAGAC AGGAATATCT ATTTTCTCC TAATATCTG 1020
TTTCAATAT GACACAGAG CTCTTTTGA ATTAAGAGT CATGTCCATC CTTCGTACA 1080
TATATGCTT TCTTTTAAA CATTCTCTT TTTTATATA ATTAATATAG TAATATAGC 1140
TATTTCTATT GAATCGAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1200
AAAAAAAA AN 1212

25

(2) INFORMATION FOR SEQ ID NO: 187:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1605 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

5 GCTTCCGAA GTTCTTTTG TCCAAATC CAGGCTTCT CTATTGTGT TCCGCGCAT 60
CCCACTCTC CTCGACCTG GAGCTTACC TTCCGAGGC CAGATCTTG CCACTCCGC 120
GCGCGGGCT ACCCGGGTT TCAGCGAGG GAGCTCTCA GCGACATGC AACTACAGCG 180
GCGCGCGCG GCGCGCGCG AATGAGCT GCGCGCGAT GCGAGGCTT CAGAGAAAC 240
ATCCAGGGG GAGCTCAGC TTATTTGAC ATGAGANCA TCGATGATC CTCAGGCTT 300
AGCTTCAGG ATTTGCGTA GCTCATGAG CCGTCGCG AGGAGAGAT AGACCTGAT 360
GCACTCATG CAGCTCTGC TGAAGAGAG GATGAGAGT TCTTGGCAT GAGGCGCTT 420
AAGGAGAC TGAAGCGCA GGTGCGAT CAGATGAG AGCCTGGAA AGACAGGCC 480
TCCAGGGCT TCACTTTGA CCGCAATC GACATCTCA GACCTACTT TATATGAG 540
CTCTCTAGG TCCAGAGG GCTCTGAG TCGATATCC CTATCAAT ATGTCACTTC 600
CCCAAGAAA TTGAGAGTA ACTATAGA CCTCATATC TGTCTCTAC TCTGTGCT 660
ATCTTACTC ATCGATGAA GAGCTCTGAC ACTATATTC GCGAGGGGAC CCGATGGGC 720

60

5 AGAGCATG GAGCTCTCT CCGCTACTG CCGAGATCT CATCTTCAAT TTACTTCTT 780
GCTTACTGT GCAAGCGCA GATCAACCTG CTCAGATGT TCGCACTGCT GCGATATGC 840
CTCTTTGGC ATGCAATGT CTTTCTATC ACTATATTA TCACTCTCA GCGCTCTTC 900
TAACCTTCT GCTTGTGAT GGTGAGACT TCCAGACTG CATTGTGAG AGTGTGTGT 960
TCTCGAGCG TGGGCGGCA AGAGGCGTG CTCTCTGAT GCAAGCTGCG TCGCTTACAC 1020
ATGCTCTTC TCTTATAT GCAATTTGCG TAACAGAA TGTATAGAG GATCTCTGAC 1080
ACATCGAGG GCTTCAGAT CCGGCGATC CAGAGGCTC CAGAGACAT CCGTGCATG 1140
CTCTCTGTC CTGCGCTTC CAGCAGCTC CTCAGGCA CAGCAGAGC TTTTCCGCTG 1200
ACCTGCACT CAGCTGAGC CCACTGAAA TTCTTGGCA GTCTCTTTC CCGAGCTGC 1260
AGAGAGAG AGACTATTA AGGAGAGTC CTATGACAT GTTTCGTAGA TCGGCTTTC 1320
AGCTGCACT GAGCTTACG TCGTATAGTA CTTCTTAT CATTGTGCG ACTTCTGAA 1380
GCGACAGGC CAGAGCTCC TCGCAGAC TCGAGGCTC TCGAGCAT CAGAGAAATG 1440
GCTCAGCTC TTTCAGAGC CTTCCGACC TACCGCTTC TTCTCTTTA TCTCTCCAC 1500
ATTTCTTTC TAATATAGA CTGTGATTT AAATGTGTA TTGAGCTG GAAAAAATA 1560
AAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAAA TCGAG 1605

(2) INFORMATION FOR SEQ ID NO: 188:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

5 ATTCGCGAT AGCGGTCAC GTGTGCTG GCGCGGGA ATGCGGCTT CAGCAGAG 60
CGGACTTCA GCGCGCGAG GCAAGCACA GAAAGCTTT ATGACTTCT ACATGAGAT 120
GAACAAATA GAGAGAGAG ACTGCTTCT ACTTCCAAA ATTCAGATT AAGAGCTAC 180
CGTCTGCTT TCTCTTACT TCAATTTGA CCAATTTGAG GTTCTTACA TAGATCTGA 240
AGTTACAGT GAGAAATTA AAGAGGCTT TCGCAATTA TCACTCTTG TCGTCTGTA 300
CAAAATCAA GATGATCTG ACAGAGACA AAGGCTTTT GAGCTGTG ACAAGCTTA 360
CAAGTTCTA CTGATCAGG ACCAAGAAA GAGGCGCTG GATGTATTC AGCGAGAAA 420
AGATATCTG GACACACTG TGAAGAGCG AAAAAAATA TTAAGAGAG AAGGAAAGC 480

60

TACATATTGA GAGGAGGATG ATTCGAGCT GTTCAGAGAA CTTGTATATA AACAGATAT 540
GAACTCTTTT GCAGAGCTGG AATATTAAAG GAAGAGAGAA AGATGATCA 600
5 AAGGAAACCA CAAAGGAG AGAGATTGA AGCTCAGAAA AAGGCCAAG GGAAGAGAA 660
GTGCGAATA AACTTTTAGG AAGTCAGAA TGTCTGTGTG GACAGCTGG GAACTTCCA 720
AGCCATAGG AAGGAGAG AAGAGAGAA AATCGAGCC TTCTGTAGAC CACCGAAGT 780
AAGATGAG CAACTGAGT GACCGCCAA GCTCAGAGC ACAGAGCTT TCCCTGCTA 840
TCTCCCTCC TCTTCGAG GACTCATCT TTCTTCCAC TTCCACCCCA ACPAGAGTA 900
15 GTATTGCTT TTATGCTAT TTCTTTTCA ATACGATTTA ATATGATCA GATTAATCT 960
TTGTATATT GAATGAGGG CTTGTGTTA AAAAAAGCC TTTCCTCTTC CCTGCCCTA 1020
GAACAGCAG TATTAGAGG TCCGACATT GGTCTGCTCT TCTCTTCCA CAGCTGTAA 1080
20 CTCAGTCTTT TCTACTTAC TGAATTTGTA TGTATTAGAA CTTCTGTGAT AGTTTGTGA 1140
AATCATCCA TTAAACATAC TCTTTAAAC AGTGTCTCTG TCACTTCAGA GACAGCTG 1200
25 GAAGGECAC CTTAGAGAG CCTTCCTCTT CAGTTCTCTG CTTCTGGGTG TCTCTCTTC 1260
GAAGGCCAG ATAGAGAGG GAACTATCT GAGCAGCAG AGCAGCATCT GATCCCTCT 1320
30 GCTCTTGGC ATGTGCCCC TGTCTACTA CCAATCATGT TGGCATAGG CCCAGCCAC 1380
CCAAAGCTTT CAGTTTCCA AGAGCTAGC GTCTCTCAC CAGTACATG TCTTAGCTG 1440
TCTCATTTG TTAGTGTAA TATCTTTAT GTATTATAAA TTTTATATAC CAAAAAATA 1500
35 AAAAAAAA ACTGCA 1516

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 681 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 189:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

60 GCTCCCATCT TCTGTGCTCT CTTATATCA CCTGTTCCTC TCCAGGAGGG GCTCAGAGC
CATCTCTCTC CTGTAGGCTCT CTGAGTCCC TCCACTTTTG GCGCTCTCAG TTATCTGCG
CAGGGGACCA TTGAGCATC CTGCTCTCT CNGGACTCAA GGTCTCAGG TATATAGCTT
55 GCGGCCCCGA TCCCTGTAAA CACTTCTCT GAGAGACTC TCAAAAGTGA CTGTATTTT
GAGTTTACCA GCAATATATC CCGACACTG AAGCAGTCTC AAGCCCHAGG ATCCAGCTT
60 CTTTGGCTCT TCTGCGCTG TCTTCCGAG ATCTTCTCTG TTGCTCATG GGAAGACTAA 360

GAGGAAAG AGAGGCTCT GCTTGGCAG CAACTGAGG GATTCATGTC CCACTGCTCC 420
TCTCTCTCC TCCCTGATAT GTGGGCTCT GCTTCCCTCTC AGGTTTGTCT GTCTCTGACC 480
5 TATCTTTACA TCCCGGAGGG GTTCTGCTCT CCTGCTCCAC CAGGTCAAGG TGTGTTCCAG 540
CAGCTTCTCT TCGGCTGCTG AATGTGTGTA CCACTGCTCC CTTTCCCTCC GCGGGGCTCA 600
10 TGTCTCTCTC CTGATGCTG CTTCTGATAT YTTTCTTTT GAAAGACTT TTAATATTA 660
AAAAAATA AAAAAACTG A 681

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1014 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 190:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

60 GCTCAGACC AGCATATGA TAAATTTCTG GAACATCAA ATTCATGTT TCTACAGCA
GTATGCTTAC AAACATCTC AGCAGACCA TCAACACGA GTCTGCTCT TTTTGTATC 120
GCTGATAT TCTCTGCCA AGTGGCTTT TTAGAGCTCT TAAAGGCCA CAAAGAGCT 180
ATCTTAGAG CACAGTCAA CAGATTTCT TATATGCA AGCCATCTG CCTCATGAT 240
30 TATATGCTAC ACTCTGAGC AACTCAGAG TCCAGAGAG CAGAGAGCA GTTGTACCA
CTGTAGCAT AGCTATTTT GCAGAACAT GTTCAGCTT TACAATATC CCTGCTTAA 360
TGAATGATA CAGAGTTCT GAACTGATG TTCAAAATG AGTGTATAAA TCCCTTCTCT 420
TCTTCTTCA ATATATCTT GAATGCGAA AAGACTACT TTATGCCCTA ACAGCTTAC 480
TTGAGATCC TTATATGAT AGAGACTTG TACACAGACA GAGCTTAT CCAATGTATAC 540
45 AGCAGATCT ACTTGGGTT TATGATTTG GTTGTGAGA TTGCTGAT CACTTGTGA 600
ACTATGATG CCGCATGTT TTGTAGCAT CTTCTCATCT AATTCAGCCA GTTATGCGAG 660
50 CCTTAGAGG CTTGAGAGT GCTATTGAC CATGTAGAT GTTCCATAT TTTTATCAG 720
GTCTCTTTCA CCGACCCCGG AAGTCTGAG ATGTATATG GAATTTTAC AACTCATCT 780
AATTTGCTC CAGAGAGCT CTAATAGAC ATTACCGAG AATCTAGCA GATATATAG 840
55 TACAGCTATA TCTGTATGA ACTGACTAT ACTTATATAT TTTATTTTAT ATTTATGCT 900
TATGAGAC TACTTACAC CTTAACTTG CTTTGTATG GTGATGTAAA CTTTAAACA 960
TTGCAATCA GTTATAGACT GTTCCATAG GGAAGAGCTA GGAATTTCTAT AGC 1014

(2) INFORMATION FOR SEQ ID NO: 191:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2779 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(141) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

5
10
15
20
25
30
35
40
45
50
55
60

TCGACGACG GTGTGTTCG ATGTGATGTC TGTGTGTGGT AGCTGTTCCT GACAGGGGAG 60
AGTTAAGCTC CCGTCTTCCA CCGTCCCGGC TGGCCAGGTC GCTTGAGGCT GACCCAGAGA 120
CCGAGACTCG CTTCCTGAG CTGATGCTGC AGAGCTGGGG AGGAGGTTTC CCGCGCTTCCT 180
CTGCTGTTCG CCGCGGAGC CCGTCCCGGC TTCACTTCTCT CCGCGAGCCC CTGCTACTCA 240
GAGCTTCGCG GATTCGACGA CCGCGGAGC CTTGCGCTCA GCTTCGGGGG CTTCAGTTCAG 300
GCGACACCG AGCGGCAATG GAGAGAGAC AGGAGCCCTG ACATCTTCAT CTGCAAGGAG 360
GTCTGTGCTG GACCGAGCAG CTCTCTCTC CTAGAGTAC CTACCCCTCC AGCTTCTCAG 420
TTTTCAGGTT GAGAGATTA GATGAGGCC AGAAGATGCG CTCTGAGCGG GACAGAGGAA 480
AGCTGAGTTT TGGAGCGGG CTGCTTCCA TGGATTCACA GTTTCAGGCC GAGAGCGGGA 540
AATTGCGCCC TCGATTAAGA GTCACTCTCA ACTACCGAAA GGGAGAGGCT CCGAGTTCAG 600
CGGATCGAAA CCGATTTCAG CGAGTTCGCG TCTTCAATGC GTTCTCCCGG GGTGTTCGCG 660
AGGATCTGCG TGGACTTCCA GATTACTTCA GCAAGACGAG CAGTAACTTC ACCGACTTCG 720
AATACACAGA GGGCTCCACA GTTAAGAGT CCGTGTGAAA GCGTGTGCTG AACCTTAAGG 780
AGGGGTCAA TGGCTCGATT CTGCACTGCG TCGAGTCCA CCGGGACTCT GCGAATCTTC 840
AGCGCCCTGG AATGCGCCAG TCGACAGATG ACTATTACCG AGCGCAGAGC GCTCTGACAA 900
TGGCCATTGA GAGAGGAGM CTGAGTGTG TGAAGCTCTT GTTGGAGATT GGGGCGCAATG 960
TGCATGCGCG GGTCTCGGCG GTTCTTCCA GAGGGGCCAA GCGACTTCTT TTTATTTCGG 1020
TGAAGTACCC CTCTTTTTCG CCGCTTCGAC CAGGCAATCG GATGTGTTAA CTAAGTCTCT 1080
GCGAGACCA CACCGCGCG CCGACTTCCA GCGCTGACT CCGAGGGCAA CAGAGTCTCG 1140
CATGCCCTAG TCAATATCT GCAAACTCA GCTGAGAAC TTGCACTGCT GACCCAGATG 1200
TATGATGGCG TCTTCGACG TGGGGCGCG CTCTGCTTCA CCGTGCAGCT TGAAGACATC 1260
CGCAACTCC AGGATCTTAC GCTCTGAG CTGGCGGCAA AGGAGGGCAA GATCGAGATT 1320
TTGAGGACAA TCTTCGACCG GAGTTTTCG GACTTCAGCC ACTTTTCGCG AAGTTTCACC 1380
GAGTGTGCT ATGGGCTCT CCGGATTCG CTGTATGAC TGGCTTCTCT GAGACAGCTCT 1440

5
10
15
20
25
30
35
40
45
50
55
60

GAGGAGACT CAGTCTGCG GATCAATGCG TTTCAATTCCA AGAGCCGCGA CCGACACCGA 1500
ATGCTGCTTT TCGAGCGCTT GAGCAAACTG CTGACGGCGA AATGGAGTCT GCTCATTCGCC 1560
AAGTCTTCTT TAAACTTCTT GTTAAATCTG ATCTACATGT TCACTCTTAC CCGTGTTCGC 1620
TACATCAGC CTACCTGTAA GAGCAGGCC GCGCTTCAGC TGAAGCGGA GGTTCGAAAC 1680
TCAATGCTGC TGAAGGCGCA CATCTTATC CTGCTAGGGG GATCTACTCT CCGTGTGGCG 1740
CAGCTGTGCT ACTTCTGGCG GCGGCAATG TTTCAATGGA TCTGTTCAT AGACAGCTAC 1800
TTTGAATTC TCTTCTCTTT CAGGCCCTG CTACAGATG TGTCCAGGT CCTGTGTTTC 1860
CTGGSCATCG AGTGTACTCT GCGCTGCTTT GTGTGTGCG TGTGTGTGGG CTGGCTGAC 1920
CTCTTTTACT ATACAGTGG CTTCAGAGAC ACAAGCAATCT ACAATGTCA TATCCAGAG 1980
CGCTGTGAG CCGTGAAGCAG GATTTGCGG CCGGAAAGCT CATTACAGGC CGAATGCGAC 2040
AGATGATGTC CAGCCCATG AGGACAGCA KGAAGAGGCG AACCGCGCC AGTACAGGGG 2100
TATCTGTGAA CCGTCTCTTG AGCTTTTCAA ATTTCATATC GGAATGCGGG AGCTGCGCTT 2160
CGAGAGAG CTGCACTTCC GCGCAATGCT GCTGTGCTG CTGCTGCGCT AGTGTGCTCT 2220
CACTACATC CTGCTGCTCA ACATGCTAT CCGCTTCATG AGCGAGCCG TCACATGCTT 2280
CGCCTATG AGCTGAGCA TCTGAAAGCT GCAAGAAAGC ATCTCTGTTC TGGAGATGGA 2340
GATGCTGAT TGTGTGTGCA GGAAGAGCA GCGGCGAGCT GTGATGCTGA CCGTTCGCGAC 2400
TAAAGCAAT GCGAGCGCCG ATGAGCGCTG GTGCTTGAGG GTGAGAGAG TAAACTGCGC 2460
TTTATGAGAG CAGAGCTGCG CTAGGCTGTG TGAAGAGCGG TCAAGGGGAG GTTGTGCTCG 2520
AACTCTGAG AAGCTGTGTC TGGCTTTCGC TCCGAGGAG GATGAGATG GTGCTCTTCA 2580
GGAAACTAT CTGCGCTTCC AGCTCTTCCA GTCCAACTCA TGGCCCAAT TGGAGAGGAG 2640
GCGAGAGAC AGAGCAGAG ATCTTTTCAA CCAATCTGCG TGGCTGTGGG GTCCCAATGA 2700
ATCTGTGCG CAATATATTA TTTTCACTTA CTCANAAAAA AAAAAAAAAA 2760
AAAAAAAAA AAAAAAGGC 2779

(2) INFORMATION FOR SEQ ID NO: 192:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1923 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(141) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

5
10
15
20
25
30
35
40
45
50
55
60

ACCCTCTCC CTGCGCTCC GCGCGCCCGT CACATGATC GCGTCGGCC
TGGCTCTGCA GCGCTCGCC TGGATCTCTCC GCGTCCTCT ACTCAGCCGC ATCCGCTTGG
ACATCATCC GCTGCGCCGC GCGCGCTGGT TGGATCTTAG GACACAGCC CAGACCTCT
GCGTCGTC GAAATCTCT CAAAGCGGCG GCGCGAGCG GTCTTACGAG GAGGCTCTC
AGAGCTCAT GAGTACGCG TGGGTATAG CAGGCGCTCC CATGCTCTTC TGTGGCTTCA
TCACTCTGGT GATCTGTTC ATCTCTCTT TCTTGGCTT CTGTGAGCC CAGATCTCTG
TCTTCTGAG AGTATGTA GGTCTCTCTT GCTTGGCTG TGTGTCTCG ATCATCTCC
TGGTATTTA CCGCTGAG TACACGGA CTTTACCTT TATGCTAC GTCGTCTCA
CTTACATTA TAACTGGCC TACGCTTTC GCTGGGAG CAGATATC CTGATCTGT
GTGCTCTT CTCTGCTCC GTCGCCACT ACAGATGA CTTTGGCC ATGCGAGC
CCAGTACT CTACACATCT GCTTACTTC GGTATCTG TGGAGAAA TCGCTGCTC
TGGATGAG TCGAGAAA GAATCTCTT CTCAGGGA CTTTGAACC ATTCTTGGC
AGTCTCTA TTATTAATCT AGTCAAAAT GCTAAATTA TTGGGAGA ATATTTTCT
AGTATGTT ATATCTTCT GTTATCTTT TATATCTTT TGTGAAGTT TGTCTTTCA
CTAATTAATCT ATACTATCC AATATTTCT TATATCTAT CATTAATTT ATACTATCT
TGTAGAGAA TATGAGCTG AACTTATGA CTTTATAG TAAATATAG GTTCTCAGA
TTTATATCT TCAATAGTT CTGTATTTT CCAATATGA TGGACTGCTT CTGTAGGG
CTAGAGAA GAGGATTA AGTATTAAG TTGTATGA CCAATATCT TAAAGAAAT
GCAAAAAA AGTTATTTT CAGGCTTC AACTATTTA GAAAGCAA ATCATTTCT
AATGATAT CATTTGTAG AATTTCTAT TATATCTCT TATATCTAT TCACTATG
GCTTCAATTT GACTGATAT GTCATCTAG AACTATCT TTCATGCTCC AATCTGTTG
CCTATGTTT TAAAGCTTTC CTTTATGCTT GAATATTTA GATGAATTT TCTCTTTTA
AGTCTTTAT AGGTTAGCG TGTGGAAA TCTATATTA ATAAATCTGT AGTGTTTGT
GTTTATAT TCGAGAGAG AGTATCTCT ATCAAGAT ATCAAGAT GACTGCTCT TATATATCA
TGAATATG ATCTGTTTAA GTTGTGAT AAGCATATG GAGGTCTAT CTGTGACAA
AGTCTCTAT AATAGCTCT CAGAGATTA ATGACTTTC TTCTTAAAT CTCAGTTTA
TCTGCTCT ATCATATGA CAGCTTTC ATGTTTGA ACTGTATGA GAACTTCA
TATATATTA ATCTGCTCT TTCTGTGTA ACAGATTTTA ATGTCTGAT ATAAATATG
CCACAGGA ATGCGGAT TTGATTTCT CTGAATGCA TATATATAT GATCTGAT
GCTATATG ATTTTATCC ATTTGACTT ACATATGA AACTATCA TTTTAAATAT

5
10
15
20
25
30
35
40
45
50
55
60

CAGATATTA TTTTATAGT TGTGAAAA GCTAATTTTA GTTTCTATTA TGAATTTTC
CCATTAACC AGTATTTCTA AAAAAAACTN GAGGCGGCGC CCGTATCCA
ATT
ATT
ATT
ATT
ATT
ATT
ATT
ATT
ATT
ATT
ATT
ATT

(1) INFORMATION FOR SEQ ID NO: 193:

(A) LENGTH: 2146 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(E) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

5
10
15
20
25
30
35
40
45
50
55
60

AGGCTCAGCG GACACTCTC AATATTAAC AGCTTTTAC AGTGGCAGA ATTGGCTTC
AGACCAAT CTGGTTCGA GTACTCATG GTGTATTCG GCAATTCCTT CCGCATCTG
GGCTTCCA TCTCTCTCT CAGTGGACA TGAAGGAC GGGGCGCCAG CAGCTGATG
GCTCAGCG ATCAAGTCTT CTCTGGCTT GGCACATTA AGAGCATCT GCTGTGGA
GCAATCTCT GGTCTCTAC CTGGATCTT GCTTGGCTG CTAACTGCTT GTCTCTTCT
CCCTATGTC CTCTGGCTG CATGACCGC CTGATCTCTT SCGCTCAGC CTCTGTCTC
AGGCTATCT TCTGATGCG ACCGCGAGC GCTTCCAGAC CAGCGAGCG GCGCATCTG
GCTCCAGAC CAGCGAGCG ACCGCGAGC GCTTCCAGAC CAGCGAGCG GCGCATCTG
ACATCACTT CACTCTGCTT GCTCTGGA AGAGGCGCG GCGCAATCTT ATGAGCTCT
CTCTAGCG GCGCTGCCA GCTGCGCGC CCGGACTCTG ATCTCTGTAG TGGCGCGCTC
CTCGCGCGC CTTTTCGC CCGCTGCGC CATCTCTCG CTAACTGCTT ATTAATCAA
AGCTTATTT GTAAATCA GCTCTGTG AGACAAATCA GTTCTATAC GTGGTGGC
TCTCCAGCG CCGGTGGCG GTGGACAAA GAGGAGAGC AAGCATCTCT GATCTGATC
CTCTTCTAT AAGCATGCG CCGTTCAGC TCTCTGCGC TCGCTCAGC ACACCAATCT
GCGAAGAA AGAGCTTAC AGCAGACTT TCAAGAGC AAGCTAGCG GCGCTGAGC
ATCTGCTTC AAGCGTCC CTGTGAGA GCGCAGCGC CCACTTCCG TTTCTCTTT
ACTGAGAG AAGCTATCC AGTTTCGGA AACAAATCT TTTTCTATTT TGGGAGCGC
GCTATATG ACATGAGC ACTCTTTTA AAGCGCAA AGAGAGCG GAAAGATCTG
GATCTATCT CAGGCTTGA GCAATTTGA ACACAAATCT TACGTATTTA ACTTGAAGA
ACGATTTT AAGTGTGCT CATCTAGAA GCTTGTATG CAGAGCAA CAGCTTCTAT
TTTCTAGA TCTCTTAT GTCCAGAA AGCAGCGAC AAATCTCTT GCGTTTACG

5 AAAAAAAAAA TTGAGCAAC GTTGGGAC ATGGTTTTTG AGGGTTTAG TTCTGCTTTC 1320
TGCCTCTCTT CCAAGCCCC AACCTCCAC CCGTCATACA TGAGCCAGTG ATATATCTTG 1380
TTGAGGAGA AGATATTTA GATTTGTTTT GATTTCTTTA GATTCAGGG CAAATTTCCA 1440
CAGCTGCTCT GCGTGTGATG AGTGTCTTGG CAGGGGCGGG AGTAGAGCA CTCGGGTGGG 1500
10 GCGGAAATG GCGTTACTG ATGTAAAGGA TTCTTTCTTG TTGTGTGTAG ATCCATGCA 1560
GTATGATTT CTGTGATCC CAGCTGTTT CAGGAATTT TGTGTGATG GCTTAATCC 1620
15 AGTTTCTAT CTTCAGAGC TGGGCTGAA GTTAACTCA GTAGCTGAC CTGTCTGACC 1680
CGTCAAGTT CTTCAGCTT CAGACTTTT TCTCTTTGTC GGGTGGGG TGCGAACTCA 1740
CGTGGAGC GGTGGCTAG AAATGTAG GATTTGAA TACATATCC ATGGGACTTT 1800
20 CATTGCTCT CTTGCTTCTT CTATTCTGTC TCCTAACCT TTGCGCGAAT GGGCAGCAC 1860
CAGTAACTT TGTGGGCGGC CAGTGGGCT GCGAGTTCC TGTACTACTG CTTGTACTTT 1920
25 TTCAATTTGG CTACCCCTGG ATTTTCTCAT AGCAAGTTTG GTCAAGTCA ATTGAATATT 1980
GTATGTAGC CAGTGGAGC CAGCAATTC TGGACCCCG CAGTTGGAG GAGGAATAG 2040
TCCAGCTTTC CAGTTGGCT CAGAGGCAAT GACTGTATG CTGCGGCCA TCACCTTGA 2100
30 GCGCTTCTCT GCGTTGACT AGAAATGTGG GGAATCGGG CAGAGAGGC TGATTAAGCA 2160
TGGAAACTA TTGTGACAA GTTCTTCAG AGCAATTTCT TAATGAGATA TTGTATTTA 2220
TTTCCAGCC ATATAAATTTG TAATTTTGA AAAAAAAAAA AAAAAAAAAA 2280
35 AAAAAAAAAA AAAAAAAAAA CAGGGGGGC CCGTACCCAA TTGCGGTAT ATGATGTAA 2340
ACATTC 2346

(2) INFORMATION FOR SEQ ID NO: 194:

- 45 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3054 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
- 50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

55 TAATCGACC ACCCTTATTT CTACATATGA TAGGCGAC TGAAATATCC TAACCCCTTA 60
AGCTGAAGGT GCGCTGTGAG AGAGCAACT GAATATAGC AGGGCTGGGC TGTGTCTTCC 120
TGGTATAGG CTGACTTTT CCGCAATTC TTCTCTGGA GCTTTGAGC CAGGTGCTA 180
60 AAGGAATG GTAGGAGC TCTTATAT ATTTCTTAA AGCAATATCT TGAAATCA 240

5 TTCAAGCTT GATGCGCTAT AACCGCTGCC TGAAATTTCTT CTAATTAGGC TATAGAACT 300
AGCAAGACT TTACATATTT CAGATGTGTT TCAATGCTTT CTAAGCTCTT CTATAGCGCC 360
CTCCATTTAT TTGACTTAAG CATACACAG TGCACTAGC ATTATACCA GAGTATGAGA 420
AATACAGTGC TTATAGGCTC TAACATTAAT GCGTTGAGTA TCAGGCTGCC CTGAGAAAG 480
GATGCGACC TGAAGGCTTC CTTATGTCTT CACCAAGAG AGCTGCTTCA TGAAGGTAT 540
CTTTTCCCC TATGCTGTTT TTGCGCTGCC CCGTCTTAAT GTTACGTGGG TACCGAGCT 600
GGTTCTTGG CTAGGTATG GGGACCAAT TCAATACCTC CTAATCAATT CTACATAGT 660
15 AAATCAAGCT ACCATGTGTA CTGCGAGAG CTGGTTTTTC CTAGTATACC CACTGCAATC 720
TACTCTTACC TGCTCAACCC CCGCTTTCCA GTTATGGGAC CTGCTAAGTG TCGAATTACC 780
TGATTAAGGA CAGGGAATTA CAGAGAGGC CTCTGTGTTT CCGTGGCTCA GCGAGCTGCC 840
CAGAGCAAT AAACCAATTA AACAAGATA CTGATCAAT TTTTATCTG GTTCTCTCTC 900
ATTCGACTG CACTGTGTC TCGTTGGCT GACTGGGAC ACCCAATAC TACAGAGTCT 960
25 GACGAGGA CTGAGACTG TCACTTTCTA GCTCGAAT TACTGTGTAA ATTAACCTTC 1020
AGAACTGCTA CCAATAGTG AAATGCGAC ATTTGCTTT ATTAATTTCTA CCAATGTGG 1080
GAAATCTGG CTTTCTTCCA GCGCTTTCCA GGGCAATAAA CTCAGCCCT TCGATAGCAA 1140
GTCCATCTAG CCAATATTT TTTTAAGAA AACTGCACT TGTTTTCTT TTACAGTTA 1200
CTCTCTCTT GCGCCAAAT TATTAATCTT AATGTATAAA AAAGTCTTA ACACAGCTT 1260
35 CTCTCTTGA AAATATGTA TTAATATCT GTATTTTAA ATTCTGCTC TGAAATATGA 1320
CTGTGCAAT CTGCACTAC TGCAATTTGG GCGTTTCCA TTGCTCTGCA TGTCTTTTAT 1380
CATTCAGGC CAGTGGACAG AGGAGAGG GAGACAGG GTCCCAACA CTTCTGTTC 1440
TTCTGAGTG ATCTGACCA AGAAGAGTA ACATGAGC GCTGCGTCC ATGCAOACT 1500
CTCCAAACA CTTATCTCTC TCGAAGTG GCGTTTCCAG GGTCTTTACT GCGAAGCT 1560
45 TAGAGCCCT CCGCAAGCTT TCGTTTTC TTCTTTACT CTTTGGCTT CAAGGAATTT 1620
TGAAAGAA ACAATATCT TTAACATCAT TTCAATTTTCA TAAATTTCCA GGGATCTG 1680
1740 AAAAAAGG CAGGTGCTT AGGCTCTG TAAAGTGA GAGAGAGAA ATCTTAAGAT
TACAGATTA AAACGATC CCAATACCA AAGCAAT AGAATGCTC TTCCATTTG 1800
CGACTTTTC TGTCTATAG AGCTATAC CTGGAGAG TAAATTTCA TTAAACCAAG 1860
55 AAATGGGTC ACTGAGCTC TGAGAGCTG ACTACTAGG CCACTCCAT CAGCTTACA 1920
GATCGAAG AGTTCGAG AGTCCAGCT CTTTAAGCT AGCTAGACA ATTAACCTG 1980
CGAATGAG CAGAGAAAT GAGAGAAAT CCACTGTGA GGTGACGCC AAGATGAAA 2040

5 GAGGAGGAG GAGGAGGAGT TGAAGGCGAG AAGAGGAGTC ATTATGATGCG GTCTGAGAGG 2100
AAGAGTCTTT GCTATTCGAG ATGTACTGCT AGTACTGTGA AGCATTTTGA GTCCGAGAT 2160
GGAGAGAGAA ATGAGCTATT GTTATATGAA TATATGCTTT TCCCTGAGGT CAGTATTTT 2220
AAGAGGTAA CTTCTATGTT TTACTGTGTT AATCTTAAA GAGAGGCGAG CTGAGGCGAT 2280
TCCCTGTAGG AGTAAAGATA AAGGATAGG AAGAGATTC AAGCTCTTAT AGAGTCAGAG 2340
CTTTCCGAGG TATTAAGCTT AATATTAGA AGTACATGAA GCGAGGCTG AATATGATCT 2400
AGTCTCTGAT AGTACCTGAC AGAGCAAGTG ATTATTAAT TTGAATATCA AACTACTTTC 2460
TTAATATCAC TTGTGCTGCC ATTTTTGCCA GGCAGGAAA TATGTCCGCC CTTACTTTC 2520
TTCTTCAA AATTAAATTC CAGCTGCCA AGATCATCTT AAGATTAAT TTGCAAGAC 2580
ATCTCTGAC CCGATGCT GTCTGAGCT CAGCCAGGT CAGCAAGAA CTTGCTGCG 2640
AAGCAAGTC CTTTACCTT CTGGGGAGG GCGATTAGT AGACTAGAG ACCAGAGTG 2700
AATGGAGAG GTGAGGACT TCACATGTT GCGCTGCGAG AGCTTATTA GAGCCAGAA 2760
CAGTGGCAG AAGGAGAGC TTGGCCGAG AAAAACTGT GGTATGCTT AATTCTGTC 2820
CAGAAATAG GTTGACAGCA AGCTTGTGG GTGATGAGG GATTTGGAG CTTGTTATCT 2880
TGTATATTC GAGCTGTGAA TTTTGTGAT GTAAAGAGCA ATATCTGTA AACTTAATCT 2940
CTGTATTAAT ATGAGGCTT ACACAGTAA AATATTCAT AAGAGTCAA AAAAAAATA 3000
AAAAACTGG AGGGGGGGCC GGTATACCAA TTATCCAAAT AAGATATGTA TTAC 3054

(2) INFORMATION FOR SEQ ID NO: 195:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 907 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 195:
GGCAGAGCT GTGGCCGAAA CTTTTCGTGC TCTGTGCTGC CAGCTACTGCG CTGGCCGCGG 60
CCCTGGCTTC GGGCTGCACC AGGCTGGGCG GGGGCTGTCCA CAGGAGCCCC CTTCCAGACA 120
GGTTCGCCAC ACCGCGCACC TTCTGGGGA AGCTGTCTGC GTTCCGCGGG ACCATATGGA 180
GGAGAGGCTT TGTGCTGACC TACAGCTGCG GTGAGCAGGG TTGCCAGGAGC GTTGTGATCC 240
TCTTGTCTCC TGTGCTGAGCA CAGGCGGCTT TTCTGTCTCC GAGTGTGGAC TTGTGATATG 300
TTGGGCTGCT GTGCTGCTCT CAGTGCCTGG CTCTCTCCCTG GGTGGAGCT TGTGTGACAA 360
GCACTGGAAA CTCTGTCTC TTGAGGCTCG GTCTGTGCTT TCCGCTCTCG GGGCTAGGCC 420

5 TGTCAAGTCT CTTGTCTCTT CCAGCTTGGA CAGCTGTGGG GCGAGCATGCG AGCCTGGCAC 480
AATCTTGAGA GGTTCAGGCT TGTGAGGCT ATGTCTGCGAG CAGTCTCTTGG GARGGCTGCT 540
CAGCAGATTC ACCTTCACTG GGAATGATGC GCTGCAGCCA GCTGGGCCCC AGGGCTTTTC 600
AGGGCAGACA CTAGAGCTT CTGGCCAGGC TGGAGCTGCT GGGGAGGCTG CTCTGGGACA 660
CTTGTGSSCG AGGGCTGCG TGAATGGTTG GGGGAGGATC CTTGCTCTTT CTTCTCTCTC 720
ATCTCTCTG CATTTCGCT TCTGTACCTG GAGCTAGGAC CAGGAGCTT TCTCTGAGCT 780
GAGTGGCTCG AGTGTCTCAT AAGCCAGAT GTTGTCTGTGG CCGAAGAAA AAAAAAATA 840
AAAAAATA AAAAAACTG GAGGGGGGGC CCGTATACCA AATGCCGGA TATGATCTTA 900
AAGATTC 907

(2) INFORMATION FOR SEQ ID NO: 196:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1290 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 196:
GCGACAGCA GCGACAGCA GTTGGCAGG GAGAGAGCA GCTTATTTCA CTAGCAGCC 60
CTCTCTGCTT CAGCAGGCT CTTGTGCTTG TGGAGAGGCT CTGAATCAG AGGGGCCCCAG 120
KAGGGCAGGA AAGAGTGGGG CAAAGCTGG CCGTCCGGCG GGTTCGGGC AGCTTTGCHA 180
TCTGAGCCA GCGCTCTCC AGGCAAGCT CTTTGAAGTT GGAATATGTA ACCGAGGCC 240
TTAGACAGG CCGTCCAGCA TGTATAGAC TTGAATCTAC TCTAAGCAA TATTTAATCC 300
AAGCTCACT AGATTGTAG TGAATCAG GACTAACCTT GAATGGGGG TTTTCCAGCC 360
TTTACGCGGA TGGCAGGC GTTCCGCGG GCGTGTGGA GCGGCTTAT CTTCTCTCT 420
TCCAGACTT GCGCTCCGAC CTCTCCGCC CCGATGGGCT GAGCCGCTCC GTTCTCTCT 480
CTGTCCATAC GTTGTAGTCC AGCTAAGAG ACAAAAGCA ACCGTTGGC CCACTGGA 540
AGGTGCTGCG AAGAGCTCC GAGCTCTCC AGTGTAGCC CTTGGATGCG CATTTCCATT 600
TGTGCTTAT TCTTCAGAA TCTGTATAG CTTGCGCTTAT AAGATATTA GCTCTCTCAT 660
GCTGTATTTA AAGGACTTTT AAGAGAAA AAAAAATA AAAAACTGA GGGGGGGGCC 720
GGTACCAAT TCGCCCATTA GTGATGTTA TTGAATTTCA CTGGCGCTC GTTTTACAA 780
CCTGTGAC TGGGAGAAC CTGGGTTTA CCAACTTTA TCGCTTTGA GCAATGCC 840

CTTTGCGCAG CTGGGTTAA TAGCGAAA NCCCGCAC CCATGCCC TTCCGACAG 900
TTTGGCGAGC CTGATGCG GAAATGGCA ATTGTAAAGG TTTAAATATT TGTATTAAAT 960
TCCGCTTAA AMTTTGTGT TAAATCTCT CAATTTTTT AACCAATTA GSCCAATTC 1020
CGCAAAATCC CCATTATTA TTCCAAAAA ATAAACCSA AANGGTTTG AATTTTTTCT 1080
TTCCCAATT TTGGAAACA AATTCGCCCT TTTTAAAAA GTTGAAACC CAACTCTCC 1140
AAGGCGAA AAGCTTTTT TTGGGGGGA AAGGGGCCC CTACTTTTA ACATCCCCC 1200
CCAACTATT TTTTGGGGG GTCCGAAAG GTCCGCCTA AACTTTTTT CGAAACCCA 1260
AAGGAAACC CCATTAAA ATTTTGGTN 1290

(2) INFORMATION FOR SEQ ID NO: 197:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

GGTGTGCTG GATGTGTG TAGGTGATT TTACGAGA TTAATGTAC AATGATGA 60
GACTCTATG GAGAAATAT TGAAGNAT TAAAGAGAC CTCATATATG GAGAAATG 120
SCTTGGAGG AATTGTATG AGCTTTACA GTATCTATT TTCACTCAA GGCATGCT 180
TTCTACACA ACTCTAATC ATAAAGGCT CTATGCAAT CTATGAATG ATAGCAAC 240
ATGCTTATG TGTATTTCT TCTTTGAGC ACTGTATTT CTACAGAAA TTTCCAGAC 300
ATTATGTAG TAGAAAAAA TCGACGAG CTGTTAAGA TCTTGATCC CATTAATATG 360
TATGTATAG TGAATCTGT AATCAATCA CTTTTTCTCT TTTATCTCTT ACCAAAAA 420
TTGTTAATT TTGCAATCA AATGTTTTA ATCTTTGAT ATTTTTAA AATCTTTTC 480
TCTCATCAT TCCCTTTTT GTGGTTTAA ATAGACTTAC TTGCATTTG AAGATGATT 540
ACTCTTCTC ATCTTACAA TATGTATAT GGTAAATTC ATAGCAATG TCAATTTGA 600
ACCAAGATT GGTGATGTT TTAAGAAA AANCTGCTT TCAATCTCT GAATATCTC 660
TTGAAATAT TCTTTTACA GGTGTAGCC AACTGAGATA CCGTATGCT GTTGATTTCT 720
TTCAATGAT CTACACTCT ATTTTACCA CTGAGCTTT TATTAATTT CTATTTTAA 780
AGTTATTTG TCTTAACTA TTTAATTAAT ATACTTTTA TCTGTTTCTG AATGGGACT 840
GAATTTTTG GATATGATA TTGATTGAA AATTTTGGG AATTTTTCT ACTTGAAAT 900
TTAGAAATCT AATGAATAT TCTAATAT ACTGAAGTA WGGTTGTGTA CAGTGACAC 960

TTCTTAATA TATGATGCT TCCCTAAAN GAGGCGGAC ATGTCCACT TTCCACAG 1020

(2) INFORMATION FOR SEQ ID NO: 198:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 524 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

AATTCGAAA GTTAGGTTT GTTGTGCTC GGGCAAGCA AGTCTTTTCA CCGACCCCTT 60
CCCGCGCAG AAGACTGAA GTTCAATTTGA GAGCTGTAT TTGGGTTTAA GCGGAGCTGC 120
TGGGGCTTT GTCCCGGCC AGGACACAG YTAATTTGAA CCGGGCGGCG CTTGGCTTAT 180
GATTTCTCT AACCGAGGG CGGCTCTGC CTTCTACTCG TCCGAGCCC ACTTGGCAGG 240
CAGAGGCTCT CCCAGCTT TCGAGGCTGC TCGAGTGCAC CTGTTGGAAAT GGACTAATAG 300
GACCTTTGT TGGCAAGG TCTCTCAAC ACCCTGCTGC TGGTGGCAG GCAAGGCCCTC 360
TGGAAAGGAA GGGCGAGAC TCAATGAGAC CTCTCTGAGC CTTGCAAGGC AGGCAATTTGG 420
CCGAGGCCA AGCAATTTGG TCTGCTTCCC CCAAGCGGAC AGAAGGCTC TTGGGGCTCT 480
TCCCTTCTG GAGAGGCCC CTTGCTTTG CTTCAATTA ACTG 524

(2) INFORMATION FOR SEQ ID NO: 199:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

GTGATACAG GAAAGTCAAT CATCATCTGT CACCAATGCA TTCTGTCCA CAGCTTTCT 60
GTGATGCCA CTTCTGCTC TTTGTGATGT CCAATATGCC CTAGCTTCT CCCCCTCTA 120
GAAGGCTTC TTGATAGAT AGAATATAG AATGATGAC ATTCTCATG TGCATATAG 180
AAGAGGCAC AAGCAATCT TTTTAAATA AAGCAAGTG TCAATCTTT TAGCTGCCA 240
ATAGAGCTT GGTCTCATCC TCTGTGAGCT AGGCTTTAA AGGCTTCTG TTTTCTCAT 300
TGTCTCAT GTTTTCCCA GGTTTTATC GG 332

(2) INFORMATION FOR SEQ ID NO: 200:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

5 CAGGAGAGC CCAAGCCCTG TCTGANTG ACATCAGTGC TTCCCTGAGC TGCCTCCCC 60
15 ACCCTCGGC ATTATCCAG GAACCTTAG TTTTCTAGAA CCAAGAGAGC TGCCTGGACT 120
CAGGAGCTGG TCCAGTAGG CTGAGTGCCA GCTCATCTCT AGAAGCTCTC TGAAGTCTG 180
20 GACTGAGAC CTGCTACTC CCAAGCCAG AGCCATCAG CCAAGCCCTC TGTAGCCAC 240
CTCCCTGTGG AGTCTGAGC TCAACCAAG CCTGCCAGC TCTGSCCTC ATTATAGGA 300
TTCTCATGAG CCAATGGGCC CTGAGGGCAG CCAATTAAAG CATCTGGCTC GTTTTTCGAA 360
25 AAAAAAAAAA AAAAA 376

(2) INFORMATION FOR SEQ ID NO: 201:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1192 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

40 CCAAGTATAT TTCTATACA TTTATTTAG TGAATTTATA ATGTTTCTTT GTATTAAAT 60
ATTAGATTAT ATCTTAGAT AATATGTTTA CTAAATTAAT AGTAAATATA TATTATTTC 120
45 AAAAAATAAAT TGTGATCTA ATGCTACCA ATTAATGTAC TTGTAGATGT ATCTTATCTT 180
AATCTGAGTC TTTCCTGCC CTAACTAGGT GTGAGGACT CTTCCTCCCT GGGGAGGTTT 240
TTCTTTTTC GAGGGAGGA GGGCTTTCC AGTAAATGTG TCTAGATGT TGGCCAGAA 300
50 AATCTGGAC CACACACAC CAGTTCTCTC CTAAATCCAC GTCAATTTGC TTCTATCCCA 360
GCTATGTTTC CAGTCTCTC TGGGTGTTTC CAGAGCCAC AAGAATAGAA TAAATCTCTG 420
55 KTGAGTTGTT TATTGTCTT TCACTGTTT TTACTGTTA WTTCTGAGT TTATGGGTGT 480
CTGTGAATTA AAAAAAATA GTGAATTA GTAAATCTCA GTTGAAGCA AATATACATA 540
AATAGATATA AACTGAGCTG TAAATATAT CAGTTTATA AACTTAGAG TTCTTATGT 600
60 TGTGTCCAAA KTTTCTCTG ATCTTTCTGA TGGGATACA AAAAAAGCAG TCAATTTGT 660

WATGTAATG GATGAGACC CGAAGAGCA GCATGCTGTG TTCTTGGGA CAGAAAGCT 720
5 TGTGTGACC AAGTCTTAC CACACCTTC ATGGAGATA GTTTATGTGC TGGACATAT 780
TTACACCGG CTTGCGAGTA AAGCTTTTA GTTTTGTGA GTGGAAAGG TCATCTTGG 840
CTAAGCAGG GGTGTTGTG CAGCGAAT GTCTATCTGC TGTAAAGCA CAGCTTCCAT 900
10 CATTATGAT GCTGCTTACT CAAAGATGT GGTCCCAAC AGCTTTTGG AGTGTCTCT 960
TGATTCAGG ATGAAGCTG GAACATCTTG AGGACTAGT TAACTATAG TCTTAAATA 1020
15 ACTCTGACA GTTTTCTCT AGTTATCTC TACATCAGG GTTGTGAGA GCTGTTCOA 1080
AGTCAATTT TCTGGGAAT ATTTCAGTG TTTATTTGA CTTTACCCA GTCTGTGTAG 1140
20 CTTTATTTCT TCTAAACTCA CCAATATCT GATTAATAT CAATTATAG GG 1192

(2) INFORMATION FOR SEQ ID NO: 202:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 589 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

ATCTTGGCT ATCTTTGACA GCGATTTCT CCAAGTTGAT GCTTCTTACA AGTGATATA 60
35 GTCTGTCCC AAGATGAG AGCTTGAAT CTACAGAT TGAATGAAT GAGAGAAACA 120
CACAGATTA GATACGAAA CCAATCTCT CAGTTCTCA TTCTAAATTT GTTCTCTGT 180
40 AATATGATG TATCTGTATG AATCAGCAC AGATGTGTA AGTACAGCTG AGTCAGATG 240
ATCACAATAC AAGGGAAT GATACAGCA CCGGAGTGA CATTAGTATT TTAGCCACTG 300
GTTCGAGGG CAGAGAGAA AGCTTAGCG AGATGTTTG TATTGTATCT ACTTGTGAT 360
45 CCGGGAATCA GCGATTTCCG TCACAGCTTA CTGATCTGA GGCATTTCT AGTGTGTAG 420
ATCAGAGCA AGCTATGAA ATTAAAGAC ACAAATCAGA GAGGGGTCTC TCAGGTTCTG 480
50 AGGTGAGCA AATCCCTGAG ACACCTTTTG AAGTCAGAG AGAGGACTC AAGAGAGAA 540
AATAGAGAG TGTTCCTGTC CACTTTCTC TCACTGAAC TCAATGCCA 589

(2) INFORMATION FOR SEQ ID NO: 203:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 847 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

5 GCGAGCGGCG CAGCGCTGCG GCGGCCGACG ACGCGTTGCG CGAGCTGGCG CTGAGAGACC 60
GAGCGCTGCG GCGAGAGGCG AACTGCGATCG TCGACATCTC CAGATGCGAC ATGATCTGCT
10 ATGAGCTGCG CCGAGATCTG AGCGGCTGAC ACGCGGCGCT GAGAGAGGAG ATGAGAGAGC 120
TGGCGGGGGA GCTGCGATGCC CTGACTGAGC TCGTTGAGAC TCGCGTGGCG CCGAGCGAGCT 180
TCCAGAGACC ACCCGGAGT CGAGGAGCT GAGGCCGACG GAGGAGAGCA GCGTACTCTTC 240
CCGAGTACTG AUTGCTGAG ATGCTCTCTG CGACTCTGCA CGAGCGCTGA CAGAGCGACT 300
CAGCTGCGAG GAGCGAGGCG GAGCTGCGCT GATGCGGTTG GCTTCTGAT GCTGCTGCGA 360
GCGAGCGCTG GCTAAGCTG GAGGCGCTTG GCGGCGCTGA GCGCGCGAGT GCGAGAGTGG 420
TCAGCGCGAC TCTGCTAC CCATCGAGAA AGACTCTGAC TATGCTGCTA TCGAGAGCT 480
CGAGCTCTCA GTTACAGT GAGGCGAGT GAGGCGAGAC TCTTGGTCC CTGCGAGAGA 540
GCGTACTAGG GCGCGCGATC CAGAGTCTG GAGGCTCTCA GTTACCGCTG GCGCGAGTGA 600
AGAGTGGT ATGAGCTG GCGCGGCGTG GAGTGGCG CCGCTGCTG GAGAGAGAG 660
30 AGAGAGACT TTTTCCGAG CTGCGAGAG CAGCTGCTG GAGGAGAGAA GTTACTCTCA 720
CGAGCTCTG CTGTTACT TGTATATAT GTTAAAGCA GAGAGAGAA AAAAAAAAA 780
AAAAAAAA 840
35

(2) INFORMATION FOR SEQ ID NO: 204:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 852 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

ACAGAGTAC TCGCGAGAG GATGCTG TCGCGCGAG CATGAGCA ACAGTGGC 60
GCGATACG GCTTCCGCA GGTGGGCTG AAGAGCGGA AGCTCCGGA ACAGTGAAC 120
TCCATGGTG AGATCTGCA GATGCGATG ATCTGTATG ACTTGCAGCA GATCTGAGC 180
AGCTGAGC GCGCGCTGGA GAGAGAGT GAGAGCTG GCGGAGCT GATGCGCTG 240
ACTGAGCTG TTAGACTG CCGCGGCG AGGAGCTTC CAGAGCGAG CAGAGCTTC 300
AGTAGCTG ACCGAGAG GAGAGAGAG GCTTCTTTC CAGTACTCA GGTGCTGAG 360

ATGCTCTCT TCGACTCTG TCGAGCAGC CCGTGAAGAA GAGCTGAG TCGAGAGAC 420
AAGCGGCGC CTGCGTTGGA GTGCGTTGCG TTGCTGATG CTGCTGAGG GAGAGCTGCG 480
5 TAAAGTGGK AGCGCTTGGC CAGCTGAGG CCGCAGGTTG GAGATGTTT ACCCGACTC 540
TCCATACCT CATGAGAGAC ACTGCTACTA TGTGCTATG GAGAGCTTC AGCTCTGAT 600
10 TACAGTGA GCGAGCTGGA GCGAGACTC CTGCGTCTCT GCGAGAGG GTACTAGGG 660
CCCGATCCA GGTCTCTGCG AGCTTCAAT TACCGCTGCG CAGCTGAG AGCTGCTAT 720
GAGCTGGCG CCGGCTGGA GGTGGCGCC CCGTGTGGA CAGAGAGG GAGAGATTT 780
15 TCCAGAGCT CAGAGAGCA CCGTGTGCG AGGAGAGT GTTACTGAG AGCTCTGCT 840
CTTACTTTG TA 852

(2) INFORMATION FOR SEQ ID NO: 205:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1154 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

GATTCGAC GAGCTTCT GAGCAGAG AGTCTCTG CCGCTGCGC ACTGCTGCT 60
CGAGAGCTG GATGCTGCT CATGCTGCT GATGAGAGA TCTCTGTT CCGCGGCTGCG 120
35 TCGCTCTCA ATGCGAGT GCGCTTCTA GAGAGCTTT GCTGATGCG CATGAGTTC 180
CTGAGCTCG GCTATGCTGA CCGTGTGCG GCGAGCATGT GCGCGAGAT GTTCTGCTG 240
40 TCGACTGAG TCACTGCTGT CTGCTGACA GCGCTGCTG TCGCTGCTG GCGCGGAGAG 300
CTGAGTTTA AGAGGGGGA GAGAGAGT GAGAGCTTC TATGAGAT CAGTATGAC 360
45 AAGAGATGA AGGATGCGC TCGCGGAGT CTACAGAG CCGCTGAT GTTACTGAT 420
ACTGAGGA AGGATGCTA TCGTGGCGC AGGATGAGC GAGAGCTGT GCGCGGCTC 480
AAGCGTTTC GCGAGTGG CCGTGAAGC GCGAGCTTC GCGAGAGT GAGCTGCTG 540
50 GTGAGATCT CAGAGTGA CATGCTGCT TATGAGCTG AGGAGATCT GAGAGCTCA 600
CAGCGGCGC TCGAGAGCA GATGAGAG CTGCGGGGA AGCTGAGT CCGTACTGAG 660
55 CTGCTHAGA CTGCGCTGCG GCGCGGCG CTTCAGAG CCGAGAGCA GTTCAAGTAG 720
CTGAGAGAG GAGAGAGAG CAGCTGCTT TCGCGAGTAC TCGAGAGT GAGCTGCTG 780
CTGAGCTG TCGAGAGC CCGTGAAG GAGCTGAG TCGAGAGAG AAGCGGCGC 840
60 CTGCGTGA GGTGCTGCG TTGCTGAG CTGCTGAGG GAGCTGCG TAAAGTGGK 900

AGGCTTGGC CCACTGAGG CCGAGGTGG GAACTGGTC ACCCGACTC TGCATACCTT 960
CATCAAAAC ACTCTACTA TGTCTGTATG GACGACTCC AGCTCTCACT TACAGTGGCA 1020
GGGCACTGA GCGAGACTC YTGCGTCCCT GGGAAAGAGG GTACTAGGGG CCGCGATCA 1080
GGATTCTGG AGCTTTCAGT TACCGCTGGC CGAGCTGAAG TACTGGGTAT GAGGCTGGGG 1140
CGGCGCTGA GGTGGCGCCC CTGTGTGGGA CACAAAGAG GACACATTT TTCCAGAGCT 1200
CGAGAGACA CTGTGTGGG AGGAGAGCT GTAACTCAAC AGCTCTGCT CTATCTTTTG 1260
TAAATATGT TAAGGCCGA AAAAAATATA AAAAAAATA AAAAAAATCG AGGGGGGCC 1320
AGACCAATC TCCATATAGT AAGKCGCCN ANUN 1354

20

(2) INFORMATION FOR SEQ ID NO: 206:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

TCCCGAGTG CAGACGAG GCGCTCTGT CTGAGGAGA ATTCAAGCT GGTGTGGAC 60
TCAGCCCTTA GNCATTCOA AGCTTATG TTGTATCAT ATCTTACGTG TTGAGAGCT 120
GACTGGAGA ACAAAATGTG CAATPACBYG AATTTTATCT TAGAGATCTG TCGAGCTAT 180
TTCTGTGACA AAGTTATAT TGTCTANTAA GAGAGTCTT AATGGCTCT CTGATATATG 240
TAATCTCAT TACAGGTGA CTTTATATG CATACATGA TTTGATGAA GAGCTGAAA 300
CAATCTGGG ATGTGTGGA AAGTTATCT TCCGCTCTT TCGTGTGTC ATTGTCTT 360
GAGAAAGGA TGGCCCTGAT GCGACGAG CCGCACTGT AATPAAAAA TAATTCACAC 420
TATCAGACTA CGAGGCACT AGACTGGA AAGACGACAG AAGACAGAG ATCCAGCCCT 480
TTCTATTTAC AGGTGACAA ACTGTGATG TGCATGTA TGTGTTTTT AAGCTGTAG 540
CACTGTACA AATGTAAAT TTGCGTAT TACGAGTGC TGTGTGATG GAGAGACAC 600
CGAGGCACT TCACTCCAG TGTGTGGC TGTCTACACC AGACAACACA GAGCTGGGT 660
CAATTTCCC TCAGTCTT ACAAAGTTC CTGAAAGGA AGTCTTTAC AAGCTGCTT 720
TCTCGATAC TGAAGGTG AGTTTCTGA ACTGCACTGA TTTTATGCA GTTGAAAAA 780
AAAAAGCT ATTCAGAGA TTTCAAGCTG TTCTAGACA TCTTCTATG GCTTTATTC 840
CTGAGAGCA ATGTTTTTAC TTTATGATA ATTCAATGTT CCGAAGANT AAGGTGAGA 900

AACGCACTT TTAATATAT AGTCTCTCT CAGAGAGACC TAAATTAGAA AGAAAAACT 960
GTGCAATTT TCAATATCTC ATCTTAAAA AACCTATATC TTAATACAA AAGTTCTTT 1020
TGAATATAG TTACACAGAA TGGCAGACG AGTTGTCTT TAAATAGATA GTGCTATAC 1080
TCATATATC GTTACTCAC TACTGCTTT AAAAAAAA ACGGCTAT TTAATGAAA 1140
CATGAGCAG GATTATAGT CCTTAACGA TATATTTTGT GACTTAAAA ATACATTTAA 1200
AATGCTCTT CTGCTTAGT ACHTGTCTA GTCCAAATGA TTAATCTAT GTACACTGA 1260
TCTTCTTCT TATTTAATA AATTATCAG AGTGAAGAA AAAAAAAA AAAAAAAA 1320
AAAAAATA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1378

(2) INFORMATION FOR SEQ ID NO: 207:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

AATCACTGC AATTAAACC CCGTCCCTC CAGAAAGTT CAGAACCGC CATGATGAC 60
CCTCATTTA GATGGGCGC AATATTNAG ATGACTGCG GCGCCGABAG ACTGACCTT 120
GAAGGGGGA CTGAGAGAA AGATCTTGA CATTCGAAA CATGTGGGC TTGTCCACA 180
CATGTATCG CTTCAATCAG AATGGGCTT TCGAGAGCA ATACTTTAT GATAGGTGG 240
ATGCTCTGA CCGTGTGTGT CATGTCTTC GTGTGCACT ACTGACATG AGCCAGCCAC 300
GCTGATGGC TGAACAGCAT TCCAGAGCC TCGAAGTGT TGTGTGTGT AAGAGAGAG 360
GGGCGCAGA GCGCGCTTT TGAATCTTT GCTGTCTGA ACTGTGAGA CACTTGGAG 420
TGATGTGCT CTATTTTCA ACTGCTCTG TTTTCTGTGA CATCTTGGAG GCGGAGCTAG 480
TCCGACACC ATCGCGGTG CTTAGGAAT GAAGAGTTC CCGGTCTGT CTCTCTCACT 540
CTGCTCTCA ATGGGGGAG GAAAGATGG CTTTGTGGC TTTTGTACA CAGCTGATGC 600
GTGCGCTGG AAGGTGTGA CAGTGGGCC TGTGTGAGG ACTGTCCAG ACGTTTACA 660
CCTTGTACC ATCAGGCTT TCTGGCTCT GATAGGTGTG ACGAAAGTG GAAGAGAG 720
GAAGAGGCT TTTTCTACA GCAATATAT TAAATAGTAG GTCAATCAC ATCTCTGTC 780
TCTTGGCCAC CTTCCCTGT GCTTCACTGA CATATAGATG ACTGACTGC AATACTTTC 840
ACCATTCCT GGAAGAGCT ACTTAGGGA ACGAGATGT AGTCTATTT CCGATACAA 900
GTAAATTTT CCACTACA GCTGGGTGT TCTTTTCT TAAATGAGC CAGTTTAT 960

5 TCCGCCGAGT GTTCACTCTT GACCTTATCT ACTGATTTTT TCTAGTGTAT ATAGAGATGG 1020
TTGGGCTTTT AAGGTTTACA GACTGTGGGC TTGGGCACTT GGCCGAGGG STTTTGTGGG 1080
GGGCTTTGCC CATTAGAAA GTACCTTTTAA GGGGCAAGA TTTGTGTAT TTCCCTTTA 1140
CACTCTTAC CTAGGGTT TTAATA 1166

10

(2) INFORMATION FOR SEQ ID NO: 208:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 697 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

5 TACTCTTAG ATTATAGA ATTACATG AGATGACATT TCCATTTGAG AAGAAATA 60
GTTCCTTTCA GTGCTTTT TTTGATCTT GGAGAGCA GACTGGGAG ACATTTGAC 120
CCAGCCCTG ATATGAGT ATCTCTAGA GCGAGGCCC AGGCAAAAC AGCAATCTA 180
GAAGTTTACA ATTGGAAT TTCTGCGAG CTTCGGAT GACTGTCAA AGCTGATCC 240
AGAACTGCC AGGTAAATC TCTCATATC TGTCTATCC ACCACTTC AGCTGCCAA 300
ATTACTAGT GCGTTGACT ATTCTCTTT ACCTTTATCA TTTAGGTGTA GCAATTCAC 360
AATACTCTC GACTTTGCA TATAGAGCT GTGTTCTCT GTGTTCCCTT GATATAGAG 420
CATCACCATT ATCTGGAAAC ATGCAATATA TCCAGATTAT TCACTCTTC CCGAGCTTC 480
CTGATTTAGA AATTGCAAG TTCTCAGGT GATCTATAC ATCTTAAAT TTGAGAACCA 540
TTGATTAAG TTATGCAAT AAGAGAGT TAGATAGGA TGGTGGCTA TCTTCTACA 600
GTTCCTCTGT TACAGAAA GTGAGATC ATTTGATGAG ACATTAAT ATTATTTCT 660
AATACTAAA AATATATA AATCTGAG GGGGGC 697

(2) INFORMATION FOR SEQ ID NO: 209:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 932 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

60 CCGTAGTAC CTCCTATAG TGGGCTGGC CCGGCGGGG GTGAGCTTGC CGAGGCTCC 60

5 GCTCCGAA ACCATTTCA AGGTAAATAA AAGTTCCTTG GGGCAGCCA GCTGAGCTT 120
GCTCACTTC AAGTCTTATG CAGCAGCAA AAGGAGTCA CTTCCGAAA ATTCCCTGA 180
GGTTATGAG CTTCCTCTCT ACTGATTTCC TGAAGTCAA TCGAGTATG TGGAGAGCC 240
AAGAGCCAG CTTCGAGAA GCATCTGCA GCTCGAGAC TATTCGAGC CATACAGAC 300
CTGTTCTAG GAAGTACT CCGAACTAA GCGCAGATG CAAGTTTGG TTCAATGGG 360
GTAGAGAGC TATGACTATC TCGAAATGC ACCTCTGGA TTTTTCGCA GACTTGTGT 420
TATTTGTTT CTGCGCTTA TTGACTCTT TTGGCTAGA GATTCAAAA TAAAGAGCT 480
AGTATATCC CTTGTTTCA TGGGATGAG TGGTCTCTC TATATCCAC AACAGCCAT 540
CGTATTCG CAGGTCAAT GCGAGATTT ATATGACTG GATTACGAG GATATATAT 600
CATAGAGAT TTTCGAGG AGACTTTCA AAGCCAGA ATGTGAGA ATTCACTGG 660
AATAGTAG AACTTCAAT GTCTCCAT CTATATAGT TATGGTAAA CATTCGAAAC 720
TCCATAGAT AATCATAT TTCTAGAGA AATGCGATA GATCTAGTA TTGATATAT 780
TAAATGCTT TTCTTTTCA GGAATACTA GACAGACT CTGTATCTT CTGTGAATC 840
ATCTAGAG CAACTAAC TGGATCTCT TCACTAGAG ATATGTACA AGCTTAGA 900
CTCTCATTC TCAATTTCT ATTATATAC CT 932

(2) INFORMATION FOR SEQ ID NO: 210:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 661 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

45 GTCTATCTT AATTAAGGCT TTCTCTTTT AAGCTTTTC AAGGAGCAG ACCACTTGA 60
AATTCCTCC TGGGTTGAT ATGTGCTAA TCAATTTAT AAAATTTAT CTGTCTTCA 120
TTTTAAGCT TTGCTATAT ATTCAGAAAT GTCTTAATA ACAAATAT TTGTATTTAA 180
TTTAGGAGG ACTAAGGGA AGAAAAACA AACTCAGTC TTTATGTAAG CTCGAAGAT 240
ATTAAGGCTT AAGGCTTTT TCTAGTTTAA TGAAGATTT TACTACTAT TTTTATATAT 300
TCTCTTTT GAGATGAACA GATCTCTGG GAATTTTGG AATTACAAAT GCAATTCAT 360
GTATCTCTC TCAAGCTCAG ATCATTTCTA TAACCCAATG ACAAATCTC TCTTTGTTT 420
ACTCTCTGT GAATGTGAG CTCAGTTTC CCAAGATCG TGTCTTTATG ATGAGTCAGA 480

540	GTGCTTTTCC TTGGGTGGAC AGTGTGCGC CCTTTAAAT TTGGTGTATG TCGTTCCGAG
600	TAATCAAAAC TCCAGTCTGA TGTGTATATG CTATCTACAC TGTAAATCT ATTAATGATT
660	ATGTTGATTA TCTTCTCTGA AGTTCTACAC TTTTCAATTT GATGAAATA AAGTTTTTTT
661	C
10	
15	(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 592 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
20	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 211:
25	GAAATGACA TTGTTAACA CACTAAACA GAAGTACTTA CTTCTGAG ATTAAATAA
30	TAATGTTGA CATGATACAT GTACATGAT GGAATGACCA GATCTTATG GTCTACATTT
35	TGCTTTATCC TGTATGATTT ACCTTCTTAT ATCTTTGTT CTTAACTGC TAAATTTCTC
40	TTCACTGTTT ATTTTCTAGT GACAGATGC TAACTATCT TACACCTCG CAGAGGGAG
45	AGAAATGCTT TTGGGTGGC GTAACTAAAT TTTTGAGTGA AATATCAATA GATGAAATG
50	GAAAGAGCCA GACAAAGA GTTATACAA AAAAACATG GTTTTTTTTG CCAATTGACT
55	GGCTCTTTAA ATATCTTACA AGCAATCTAC GTTAAACATC ACTTTTATG AATAAATAG
60	TGCAATCTTA GTATGTGCTT CAAGAAGCCA AATGTGCTTT AGTGGCTTAA GCGTAATTTT
65	TGTCATTTG ACATCAGAGA TGTGTATAGT ATTGACCTTA ATACGCACT ATTTCTCAT
70	AGTGAATTTT TTTTGGCTAG CATTTCCTTT ACCACTAACC TTGTGTGATA GC

240 CCGCAGCTGG ACCGCGATG TGCATTAT GCGCTCGAG AAAAATCTG AATCGGAAA
 300 TGACACTAGG AATGTAACT CCGACATT TATCCCAT TATCGCTGT TTTCTTAATT
 5 TTGTTCTTG TCGCGAATG TGGCTTTCA ATTAAATAT CTTTCTCTG TTTCGAACTT
 360 TTCTTTTGA CTGTATAT TATGGTTTG GAAAGATCA TAAATATGAG AGAGTGTTC
 420 AACCGAGA TACAAGAG TCTCATAG TAACTGTAT AATCTGTAT ATGCTTTT ACAGCGCT
 480 ACATTAGG ATGTATAT TACGAAAT ATATCTCT GTATCTCT CTACTCAAT
 540 AATGATCT CTCGATAT TCGCTCTG TCCATCTGG CAATTACTGG CCAATCTCT
 600 TTATGAGAG CAGCAGAA CCAATGAT GTCTGCTT CAGAAATCT CTCTCATTT
 660 TCTCGAAC TTATTTCT AATCTCTG TTTCTTTGA TTTGTGTAT CTCACAGAC
 720 CATTTGACC ATGTGATGC CTCATTGCA TGAATGTAT CTGTGTTTA ATGTGAAAT
 780 CATTTTCA TTGAAGCTG ATGACTGCT AGCGTTTAT TTTTCTGTA AGCTCATGT
 840 CTGGAACCA ACCGAGCT TTAATACT GTGTAGAGA AAACCGAAA ATCTGTGTG
 900 GTGTGCTTT CCGTGTCAA CTCATAAA ATTCTTT
 938

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1079 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 213:

540	GTGCTTTTCC TTGGGTGGAC AGTGTGCGC CCTTTAAAT TTGGTGTATG TCGTTCCGAG
600	TAATCAAAAC TCCAGTCTGA TGTGTATATG CTATCTACAC TGTAAATCT ATTAATGATT
660	ATGTTGATTA TCTTCTCTGA AGTTCTACAC TTTTCAATTT GATGAAATA AAGTTTTTTT
661	C
10	
15	(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 592 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
20	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 211:
25	GAAATGACA TTGTTAACA CACTAAACA GAAGTACTTA CTTCTGAG ATTAAATAA
30	TAATGTTGA CATGATACAT GTACATGAT GGAATGACCA GATCTTATG GTCTACATTT
35	TGCTTTATCC TGTATGATTT ACCTTCTTAT ATCTTTGTT CTTAACTGC TAAATTTCTC
40	TTCACTGTTT ATTTTCTAGT GACAGATGC TAACTATCT TACACCTCG CAGAGGGAG
45	AGAAATGCTT TTGGGTGGC GTAACTAAAT TTTTGAGTGA AATATCAATA GATGAAATG
50	GAAAGAGCCA GACAAAGA GTTATACAA AAAAACATG GTTTTTTTTG CCAATTTGACT
55	GCTCTTTTAA ATATCTTACA AGCAATCTAC GTTAAACATC ACTTTTATG AATAAATAG
60	TGCAATCTTA GTATGTGCTT CAAGAAGCCA AATGTGCTTT AGTGGCTTAA GCTAAATTT
65	TGTCATTTG ACATCAGAGA TGTGTATGAT ATTGCATTT ATACGCACT ATTTCTCAAT
70	AGTAAATTT TTTTGGCTAG CATTTCATTT ACCACTAAC TTTGTGATA GC

TCAGCCGCTA CCGTACAGAG ATCTCTTCC AGATCACTGA TCCCTCTGGC ACCTCACTGA 720
CCACCAAT CCGCAGGCTC ATCAAGACA CCGTCTGCTT CTGAGGCTGC CTGAGCTCT 760
CGAGCTCTT TCACTGCTC CAGCTCTGC CTTTGGGAA TTGCTCTTTT GCGCTCTTG 840
GCTCTGGAC CTGCTCTGG TCACTGCTA GACTTGGAG GCGAGGCTC CCGTCTGCTC 900
TTGCTTTCT GGTCTGAGC CTGAGCTCTT CCGTTTATC TTGCTCTGAG GTTCACTCT 960
GCTCTACTG TCTCTGCTA GCGCTGCTG GGTCTCTCTT CTTTCTGAC TGTACAGAG 1020
AGCCACACT CCGATGGGA ATAAATTTCA GACCTGAGT TTGCTCTGA AAAAAAA 1079

5
10
15
20
25

(1) INFORMATION FOR SEQ ID NO: 214:
(A) LENGTH: 3791 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

30
35
40
45
50
55
60

TCAGCTGCT GATGGGCT GCGAGAGGG ATTGTGTCTT CATTAATCA AGAGTGAAG 1020
GGCTATGCT GAAGATTGG TTATGGACA TCAATTCGG AGCCAGCAA ATGATATAC 1080
5 GTCTGAGCT GAGATCAAT TTGGAGACT TGGCGGCTCA GAGCTGGCG GAGGCGGAG 1140
ACGAGTGA CAGTGGGCTG GTGGGCGCC AACCGTGGC AGCAGGAGC ACAGTCAAG 1200
10 TCGTCTGCT CTGATGTGG ATGACCCAGA GCGATTGCC GCTCTGCTT AACTGATGC 1260
CATTAAGCA CCGTCTGCTC TTCTGACC CTCTGTCTCA AGCTTTTGC ATCTTTAGG 1320
ATTCCAGCG ACTAGAAAT TAAAAAATA AGAGCTGTCA TTCTATCAT TCACACTTA 1380
15 AGCTGATTT TTATCTGTTT TAAAAATGA CTCTGCGCC TACAGAGAG TAACAAATAT 1440
GCTATGCT TTCTATTTA GAATGTATT GGTAGAGGG ATGTTTTCA ATTTTTCAG 1500
GATTAATCA TCTTCAGAA TACTTTTCTA TTCTGCTTG CAATATGCA TTTCMAACT 1560
20 TCAATATAG GTGTGACAG TGTATACCG TTTAAGCTT TCACCTCAT TGTGTTTTT 1620
ATTATGAT TTAGAGTTC CCGCATTAC AACTGTGTTT TAATATGCG AATATCTGT 1680
25 TTAAATACT GCTTTGATA TTCAAGAG GTCACTGGG ACATGTTAA CTTTCAATTT 1740
TCAATTTTA TCTGTGTGG AATCTACT ATATGTATTT TACTATGTT TTATATTTT 1800
CATTTTGGG GAAAAATCT TTTCATCTC TCAATAGCG TGTATATAT ATATCTTAA 1860
30 TCTTATATA CAGAAATAC ACTACTGAA CAATTTCAA GCAATTTGG TTATATACC 1920
CTTCTCTCT GCATGCTCA TTAGTTTCA ATTAATAGT ATTATATTT TCACTATAT 1980
35 TTACTTTTT ATGCTTGAG TTCTCCATTT TAAATCTTA ACTAGATAT TTAAATCTG 2040
AAGTTCTTT AACTACTTA TTGTGTGAG GCAATCTGT TCAATGAG TAGTTTATA 2100
GZATGCGTT TTTCCTGCC CTTCACAGG GTGGGTGGA TAACTTATTT TGGCCATGT 2160
40 GTATATTTA ACTGTCTG TAAATATAT GTCTGGCAT TTGTATGAT TTCTGTGTGT 2220
GAAGCTGCC AATATCAA TGTATCTCC ATATATAGC ACCATTTAC CTTCTCTGT 2280
45 TCTAAGCAA AACCAAGG GCGCTGTGG GTAGGTGAG GTGGGGAGT ATTTTAAAT 2340
TTGGATTTG GAGAGAGAC AGCTTTACTT TGTAGGTTG GAGCAGGAG ACTATACAT 2400
50 AATATTAAC CAATACTTT TACTGTCTT AATTTCTTA GATGTATAT ATTGTGTGT 2460
AAGTTGAGT TTCCAGAAA ACTGTATAT ATCTCTCTC TCTCTCTCA TTGAAATAT 2520
AGGTAAATA TGAATCTTA TATGGAGC ATCACTACT ATTAAACAC ACATAGATG 2580
55 ATCAATTA AAGTTTTCT AGATGTCT TTATTTCTG CACATTTAT GATTAACAT 2640
GAAGATTT TAAAAATTT TTAAAGAT GTTGTGAG TCAATTTTAT AATGTCTTA 2700
60 CCGTATATG GTATGTCCA GTTTTAAAA TATGGAGAT CTTCAATCTT AACATTTCT 2760

ATTAGCTGA TTGCTTCCA CATATCTTC TAAAGCAAC TTATTATGTTA TAAGGTATAC 2820
TTTTTGATA AGATTATTA ATCTAGTTA CTTACTATTC TGCATTTTA GAGAGAGGT 2880
5 AATTGTTTT AATGATGAT AACTGTGTC TGTGTGTTG GATCTATGTA TGTGAGCAT 2940
GTTCGACT GGTGTATG TCTATTTAA TTATTATTT ACACATATAC GTCTATCCA 3000
GAGATTAT TAGTCCATAT GATCTATGA CCAATGTTC ATTGAGACAG CAACATACCC 3060
ACTCTAAT CAGTGTGTT AGACTTTTCA AATATCTTAC TCAATTTCAA ACATATACCA 3120
TGTATTAAA ACCCTTGAT TTCCAGCAC ATACTATGA AAGCACTTC TACTCAAAAC 3180
15 ACATCTTC ATGTCTATCC ATTCTGTG TGAGAGCAA CATAGCAATA TCTGTATAT 3240
TGCAGCTTT CAGATACCC TGAATTTAAA AAGTTGTGTC ATTAGTTGTA TCTGATGAT 3300
ATAAATTC CTTCTAGTTC ACTTTGTGTC AAGAGCTAAA ACTGTGAAC TAACTTTCTC 3360
20 TTATGTGTC GTATATATG AATATTAAGA TTATTTTCA TGTCTCTTC TTAATATGTA 3420
TAAATCAAT CAATATGAT CATCTTATTT GTCATGTGTT TCTGTGKTC TCACTGTGT 3480
25 GCACACCTT GTGTGTTTAT AATTTTAAA TTGATTTTA TATGGGGTTT TTATTTGCTA 3540
AATACACGC TTTTGATCA CATTTGAAA GGGTACTTAT CTATATGACT AATGACTTAA 3600
TTGGGAATG TGAATCTTG TAAATATCAA AATCCAGGA CTTCTTGGTA TTTATTTAA 3660
30 TTGTACTTC TTAGAGAGA TACATTTT TTGATATG AAGTTTAC ATACCAATG 3720
CTACTTTTG TTGACACG GGGTATATG TCCGGGGGA AATCCCTTAC AGTTATGAT 3780
35 CCCAGTATG G 3791

(2) INFORMATION FOR SEQ ID NO: 215:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1334 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

CAGTGTGTC TCTGTCTGG GCGCTTGGG CCGCGGGGTT CCGCATGACC AGTGAAGTGG 60
50 ACATCTTCTT GGGGACAGA CCGTATGCA GAGAGAGTGT TATGGCTCTT GCTGTGATGG 120
TTACTGTGTT ACCGACGCGG TGGCGCTTGG GGTGGGCTGG GAAATCTCTG AGCGACTGG 180
55 GCGCAGGCGA CCGGTCTTAC AGAGGACAC CATGGACCAT TACCGCACTT TCCCATGCT 240
CGAGGGGCTG CTGCAATGCG GCGCAGAGT ACTGACACAG TCAATCTTCC AGATTCGCCC 300
60 CTCGCGGACG GCACTACTCA TCGAGAGGTA CTATGCTTTT RATGAGGCTT TTGTTTGGGA 360

GGTGTGGCG AAGAGCTCT CAAAGGCAC CAGAGAGAC CTGGATGACA TCAAGAGGAA 420
5 AACAGGCATC ACCCTCAAGA GCTGGCGGAG ACAATTGAC AACTTTTAAAC GGTCTTTGAA 480
GGTGGTAGAG GAATTCGCGG GCTGCTGCTT GCAATATAT CAGCAAGACT TCTCTCTCTC 540
TAAAGGTTTG GCGAGGGAAT ATGAGGCAT GCTCTCTTTT CATTACAAAC GCTTTGAGAC 600
10 AGCGAAGAAA AACTGTCAAT ATCTGAGCTT CGTTGACTTT GCTTCTCTGG CTGAGCTCAT 660
GATCCAAAC TGGACCCCTT GAGCGCTTGA CTCACAGATG GATGACATG ACATGAGCTT 720
15 AAGCAAGAAA TTCTTCCAGC ACTTGAAGGA GCTCAAGGTTG CTATGTCCTG ACAGAGACT 780
TCTGAGCTG CAGAGAGCC TGGTGTGAC TGTCTCTCTG GAAAGCTGCG GGTCTCTCTC 840
TGAATGAAA GCGAATCTTA AAGACTCTC CCGGGGGCTG GTGAAGCTG CCGCCAGCT 900
20 GAGCCCAAT AAGATGTCA GAGACTCTT TGTGACTCT TGTGAGAT TTGTGGAAC 960
CTGCGCTCC GACACTGCG CACTGAAGGA GGTGGGTTTC TTCTGTAATC AGTATTCAG 1020
25 GTCTGTGAC TGGCTGATG GCTTTCCACA CAGAGCTCT GAGAGCTGTA CATGGGAC 1080
CTCGGGGCT GCTCTCTGG CCGTATATAT GACTGAAGTT CTTTCCAAAG CTGCGGCCAC 1140
GCTGACATA AAGTTGCTCT GATTTTGGAG ACTGATCTC GCTCCCGGGA GCAAGTGGGG 1200
30 GCGGTGAGA TGTGCTGTC TGTGCTCTG AGCACTGCT GTCCGTGTAC AAGATGATAT 1260
GTATNCTG GCTCTCTGG AACTGAGACA TATCTGAGG AATGCTGCT GTGCTGAGCC 1320
35 CATCCACAG AAGA 1334

(2) INFORMATION FOR SEQ ID NO: 216:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1511 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

GTGGGGGGA TCTGTGAGG GGTCTCTCT CCGCAGCGGG GCGCGCTGCG TACCTCTCAG 60
50 ACTGTGCGCT ATGCTTCAA GCTGTTACC GCGCAGCTC GTGTGATGCA CTTTCAAGCG 120
CAGAGCTGA TGGCTGTGAC TGAATATAT CCGCCCAAAC CAGCATTCCA CCGATCTCC 180
55 CTGCAATCTC CTGCAAGCC CCGCAAGAG GAGATAGGCG TCACTAGGCT TCTCCGCGCG 240
GAGATAGCA CAGTTTCCA GAGACCGA ATGATAGCG TCTGCCAGAA TGTGCTCTG 300
60 ATGCGAGAG AAGAGCTCT TATGCGAAC CAGCTGCGGA AACAGAT CTTGATGAG 360

5
10
15
20
25
30
35
40

RTCTTCCCA ACCAGCTCTT GAGGCTCTT CTGGAGATT CCAGTACCA AATCTGCTG
CCCCCTTTTG TGGGGACAA CATCTCTGT GTAGTGAAG AGCCAGAGT CAGAGAGATG
GTACAGATCT TAGAGACTGT GCAATCTGT CCGCTCTGAG GTGGCTGAT TGTATGACCC
ATCTCTGACA GGCAGGGCTT TATGACTAC TCCAGCTCC CCAGCTTCC CCGTCTGAG
GGGAGCTTG TAGAGGCTT CACTCTGCT ACAGCCGACA CCACTCTCT GCTCAGCAC
CAGCCCCCTC AGCTGACAC CCTCTTGAC CAGTACATCA GAGAGCAGG CAGAGAGGAT
TCTGTCTAT GTGCATATG GAGGCCATG CCTGACACTG TTCCGACATC TTAGCCAGCC
TGTTTAGCCA GCGCTGGCA TAAATACAT CTGCTTATT GCTCTGTCT TCTCTAATGG
GACATGTGCA AGAATCTGG GTGCGGAT GTTATTGTCA CTTGCTTTT ACTAGTATG
ATATGTGAG GTATAGGCC ACTTGAGAT GCAAGAGAT CCATTTGAGA TGTAGTAC
CGCTCTCTC CTTAGTTTTC CAACTTGGG ACCTGATAG AGCAAGTCT CTCATCTTC
CAGTCTCAG CGAGATATC TGAAGATTA GGGCTATTGT CCCCCTGCTC CTTGTCTACT
GCTCTCTCT CAGGGGCTC CTGAGCCAC CCGTTGGGG ACMACTTCC ACTGCCAG
TAGCTCACC ACGAGTAT CTGAGATG GCACTGTGT AGAGCTCTT GTGTGCGAG
CTTTCTGCT AGTCTGTAT ATGTATAT TCTTTTACT GTAGCCATCT TTACCCATT
TGCAGAGAA GAGAGCAGA AATTATGTG CTTGCTCAG GTCATCATG TAGTATGTG
CAGACAGGG ACTGACCA AGCCCTCTG CTGAGAGCC GCGTCTGAA TTCTTCTACT
AGAGCTCTT CATCAGTTA CCGAGATG GTTCCATCC ACATCTCAG TGTCTTCCA
TGTATCTCT CAGCTTCCA GTGTACCT GTGAAAGTT TGGGCACT GCTTATAT
AATATGAAAT A

(2) INFORMATION FOR SEQ ID NO: 217:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 642 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

55
60

AGGCTTACT TTCTCTCCA CAGAGATC GAGGCAACC TAGCTCTGAC TTGCACTGT
GACAAATTC AGTACAGG TCTAGCCAA GACTGGCCA TTGAGCAG GAGAGGACC
TGTGATCTG ACCTGAGCC GGGCCCTTC ACCTTGGCTG GCTGTCTCT GTCTCTTAG
TTTGTCTAG TTGCTCTGT TTGATCTCT CACTAGCTG ATAGCCTG GAGGGGATG

5
10
15

ACCCGCTTG GAGTGTTC TTAACTCA TCCATATAT AGGGCCCTGG CATGTGTCTA
GAGTAAAGC AGATGATGG TGTTTTAGA CCAGGCTTG GAGCAGGCC TCTACACT
AATTTCTCT CCTGTAGCT GACAAAGT CTAAATAGC TTACAAAG AGAGGCTGC
CGTAGCCAG AGTCTGAGG GCAATCTTT CAGTTTCTCT TGTGTAGAT TCTCTCTAG
TTCTATGAA AGCAGAGC CTTAGGGGC CTGGCAGAG AACACACA TCTTAGGCT
GAGTGTGAA CAGAGGGGG TTGTGTCTCT GTTCTGTTC TGTCTTCC GAACTTCTC
AATMAAGCT AATCTTATT TTATATTAC GTAGTGTCTG GG

(2) INFORMATION FOR SEQ ID NO: 218:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1241 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

30
35
40
45
50
55
60

GTCTCCACT TTCCATTTTA TCTATATGA TCTCATCTTA GGGCCAGCC GTCTCTGAC
TGATGTGTT CCGTTTACC CTGGCATGT ATATAGATAT TTGTGTGAT GAGAGACC
AATAGGCCA GATAGTCCC CAGGCCCTG ATATCATTA MAGCTTGGG AATCATATAT
GTAAATGTC TTAGTCTTTT TGTGTTTTA GAAAAAATA ACAAGATGG CTCAGATGA
TCTTACCTA AATATGCTC CTAGCTGTGT ACTCATACT TTCTTTTGA TTGATATGT
AAGGAGCA GAGGAAAGG AATTAATG TCTCTTAGT ATCTCTGCA CTCAGTCTG
ACTATGCGA TATTAAGCTA TATGAAATG CAGAGATTO TATCTGAAC AGGAGACAG
TTTGACAT TTATCATGC TTCTATTAC ATATTAAGT AAACCATTA ATAAACATAT
GAATATCCA TTGCAGAG CAAAGCAC TAAACCTTTT GTTCTTTTT CTACATAGA
GAATGAT TTATTTTAT TTTTATGGG GAACTATAT AATATGACC CAGTATGTC
TTTTGTGAC TTAGCTTAT GATTCAGT TACATGAG TTGATCTGAG ATGTTACTA
CCTGAAAG CATGTGTGT CTTATGTGA CAGAGCCAG AGCTGCTGG GAATAACAA
AGCAATTA TCCAGAAC AACTCTAGC TTATGTGCA CATGGAGTG GTACAGACT
CCCAATGT GGGCTTTGG ATTTCCACAC CATCCAGT GTGTCTCATC TTCTCTTTC
ACCTCTTGA TGAATTTG AATATGTA AATGACTCT GAATTTGCT ATAGCATGAG
CAGTTCTTA TGACATATA ACAAATGTT CATATGTGA ATATTAGAA CTGTTACAC

CTCCAGTTAC CATATTTTC CATTTTGTG GATTCATAT TCATATACA GGCCTAGCA 1020
ATTACTGCA AGTTTAGCC TTGGGTAT ACCTTAGGT TATTTAATA TTGTAAATTT 1080
TATTTAATG TTCTGATG TTGCAAGCA ACAAATAT CAGCATGCC TCTTTCCAT 1140
GGCTTTAT TTACCCCT TTTCCTAAG AAAAAGAAC ATGTCTTAA TATATTTTA 1200
AAGTTTGG TATAGTTTCT ATTTCCAT TTATTAAG T 1241

(2) INFORMATION FOR SEQ ID NO: 219:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

15 TTTTATATG ACTTAACA TACACATG CACACACA TACATCCA TCAATCAT 60
CATTAAGTG GTGTTCCAG TCTCTGTG TCACTGTTA TCCAGTTTC ATTTCCAGT 120
GAATATGAG TCGAGGCA CTTTCTAC CAGATGCT TTTCAGACA AAGACGGGG 180
30 TATTAAGAA GAGTTGAA AGAGGAGAG GCAAGGAAG AGGCTTTAA ATTGAAGT 240
TAATTTCTA ACAGGAAGT GGGCTGAT ACTACATGT TATACCTCC ATCTTTGCA 300
GTTGGGATG GAACATGCT TGTATGTC TGTGAGGCT ATAAATCAT ATATCCAAA 360
35 AACACACAT CCATCATCA ACATACAT GGTTCGGAT GACAGCTCA ATAGTTTGA 420
GAGGAGTTT GTTCTTTT TTTTCTAT ATACTCTTA ATGTTGTCA GTTATCAAC 480
40 AAACAAACG AAAAAATG TTGGAAAA CTTGCATAG CTTTCTCT ATAGTCTTT 540
AAATATAGA CTAAATAC ACATCTGCC AGTTTCTCT TACATGACA GTATGCTTAC 600
CTGCCATTA ATATTAGCT GTATTTTC TCAGTATAT TTACCTGCA CTGTATTTG 660
45 TATTTTAA ACBAAAAA ACATCAAA AAGAAAAAT TACTGTAG GCTTCATAT 720
ACTATATAT TATATATAT ATTTGACAT TTTCGATAC TGTGAATAT TATCTCTGC 780
50 ATATACTTTA TACGAAGTA TTACGCTTA AAAATACAA ATTAATTTT ACAGGTTTC 840
TGTTTTGCT GGAAGTAA TTGATGTGC TAAGATCAT GTTTGTTTT TTGGGTTTT 900
TGTGTTTTT TTTTAAATG TTACAGCA TTTTTTTGA AGTTTCACT TCCGAGTAT 960
55 TGTACAGTT CACATGTT GTGATTTG ATATGAGG ATATTTAA AAAAAAAA 1020
AAACACGG GGGGGCCGG TCCATGAG CCAGAGGGG GGTTCAGGG GTACAGGGG 1080

60

(2) INFORMATION FOR SEQ ID NO: 220:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1258 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

10 TGAATTGAG GCTTAAGAT AACATATGG GRTTGAGTT GTGTGTGAT AGGTTTTC 60
15 TCCATTTTG ATTGAAGTT ATCCATATA ATTTTTCAT GTAAATTTT ATTAAGTAT 120
AATTACATA TATTTCAAT GATTTTGTCT CTGAGGTTT TCAATGTCT TCACTAGAG 180
20 ATAGTGGGC TTCAAGCTG GATTAATGCA GGGCAGGAG GCTTCACTC TATGACAGA 240
TTGTATTAT GGAATGCTC TCCCTGTCT ATATGTTTG ACOTTAACA TCCACTTAC 300
TTGACACA GCGAAGGTT GAACAGGAC CTAGACAGCA AGCTCAACT ACCCAATCA 360
25 GAGCGGTGC CTTGCTGCT TTGGCCAC AATGTGATC TGTCCCTTG GCGATGAGC 420
CGGAGAGCA TTGACCGTT CAGTAAGAG AACGTTTCA CAGGTTGGC AGAAATCA 480
30 GTACAGCA ACAAAATAT TAATGAGCT ATGAGAGTC TCAATTGAA GATGATGAG 540
AATTCAGAG AAGATATAT GTCTTTTCC ACCCAAGGG ACTACATCA TCTACAAAC 600
AAGTCTCCA CTTGCTGCT CTGCTAGT TTTTGTGTT ATTTCCATC CAGTCTGCG 660
35 CAGCTCTTT AGTCTCTC CTTTGGTTT CCGACTGAC TAATTTTAT ATGATATTT 720
AATGTCTCC TCACTACT CCAATAGA GCGCATTTT CACTTAGAA AGACACTCG 780
AACCGGTT CATTTCTCA TCTCTGAT TACGCTTGA CATTTGCTG TCTACATTA 840
40 GTCCAGTTA GTCCCTGCG TGTAGATCT TCTCATCAG CTTCAATTT TCAATCGGA 900
TTTGTGAGA AGCATGAA ATCAGACT CCGTTTAGA GATGTAAT CTCACTACT 960
45 TCTGAGCTT TTTTCCAT TCAATTTAT TCAATCATG ACTTCCATT CAGAGGAAA 1020
TGAGTCAAA TGTATTTCC CAATTTCT GTAGCGGTT GTTTCAGAT CTTCCTGCT 1080
50 TCGAATGAA ACATGAT CTGATGCA GAAGAGGGG TCTGGGATC TGTGATTTT 1140
TCCCTACTAG AGTGTCCA GAGTCACTG TATTTTGA ACTTCTAAC TCAATATTA 1200
GTTCCTCTG TCTTGGCAT CAGATATAT TCAATTTT TGGGCGGGG CAGGTTGG 1258

(2) INFORMATION FOR SEQ ID NO: 221:

(1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1693 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

CACATATAT GAATATATAC CTTCTAATAA AGAGAAAAAA AATATCAGGC GGTCAACTT 60
AGAGCAACTT TGTCTATATA AGCAATAGTT TATTTCACTA GAATAAAATTT AATATCAAGG 120
ACTATATCAT ACTCAATATG TAGGAGATTC TTTTAAATAT GACTCTTAAA ACAATCACTG 180
AATCTTGAAT CACATATACA CCGTGTATAT TTTCTTTAAA CACTCTGAAA GCTTATGCTT 240
CTGCAATATCA TGTGCAACTT GTGATGAGCA GTAAATATAC AGAGAAGATG TTTATATAGCA 300
ATTAAAGGCT GTTTGCACTT TTAAAGACA CCGTGGGCTGT AATGATTTCT GGGGGCAGAG 360
TGGCATATAT TTTTATAAA AATATGACAT ATGTCAATAT TTTGCAATGT TGTATTCCTTG 420
TTCAATTTTAT GAACAGCAG TTGACCAATC ATAGAAAGATA TTACTTTCTT TCAATATGCTT 480
TTTGTCTTAC TGGCTTATGA GCTTTCTTATG AATATCTATG GGCACAGCAG CATATCAAGTT 540
TCCATCTCTTAA TAGCAATATGA ATTAATTTTAT TATCTACTGA TACAGCAATC TGGGTCTCAAT 600
GAATAAAAT CATTTTATCC GTCTTTTATG TATATGTTTAA AATATATATAT TATATGTCTT 660
GCAATATGCA GAACAGCTCT GAGAGCAACA GTTTCCCAAT TACTCTTTCT GACCAATATG 720
GCTGCACTAG TTGCTTCTCT TTGCGAAGA GGAAGAGGATG TGTGAACATG GCTTACATATC 780
TTCCATATAC CAATATATGA TACATATAT TACATATTTA TTTTATGCTT CAGTATATTA 840
TTATTTAATTT TTTTATGTTA TGGCTATCTC TTGCTCTATTT AAGAGAAAG CAATATCAAT 900
GAGATATCAG CATATGTTTATG TTTAATATCT TCCAGATTTAC ATGTTTTTATC AGTGGCTTCC 960
TATTTAGGAAA AGATATTTCTT CTATATCACT TGTTTTATAC TTTTGAACAA TTGACAGAAA 1020
TTATGCAATG GTTTGTTTATG ATAGCACTT GATGCTCTG TTTTATCAAT TTTGCTTCAA 1080
ATGCGCTTAC TCAAGAGCC CTAAAGCTG TAGAAATTTA AAGATTTTCA AAAACTTTCTT 1140
ATTCTCTTCT TAACTCTACC ACAAATCTAG GATTTGTGATA GCAATGATG GTATATGAAA 1200
GAATATTTTCA CCAATTTTCT TTTTATTTTG TTTTATTTCT TTTGATTTTG AATATCAATCT 1260
TATTTCTTTT AAGATTTTCT ATGATATGAT GTGAGATGTC TAGCGAATCT ATGCTCTGAT 1320
ATTATCTGTA AGTCTCTCTT CACCTCTTAC AAGATTTTAC ATGCTCTACT CATATATATG 1380
ATTCTCTTAT TAGATATGAT TTAATATATC AATGATCTTT TAAAGAGATG ATGCAATCTT 1440
GTATTTATATG TGGCTCTCTT GGTCTTATCT GAGGCAATTT AATCAAGTTT CATATATAT 1500
TTTCCAGTGT TGAATCTCAC AACTGTACTT TTAATAATTT CTTTCAATCC TGAATATACA 1560
ATAGAGAGG CCAATATAT TGGCTCTCTTA TCTTGAAT TTTCACTACT TTATTTTAAA 1620

AGTTGTGTAT AATTTCTTAA ATCTGTGAAA GAATTAATAAG TGAATTTTAA TTAAAAAAA 1680

5

AAAAAAAAA AAA

1693

(2) INFORMATION FOR SEQ ID NO: 222:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1196 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

AGCGTGGGT CGACCCAGGC GTCCGCACTT TGGCTGTGTG GGAAGGGAG AAGATTTGT 60
AAACCCCGCA CCGAGTTCTT GCTTACCGCA GCGCCGCTCT GTCCGAGAC CCCCAGGTCA 120
AGCCACCTTC ATCATGTCTG ACCAGAGGC AAAAACTTCA ACTGAGACT TGGGGATTA 180
GAAGAGGCT GATATATTA AACTCAAGT CATTTGACAG GATAGAGTG AGATCACTT 240
CAATGTAAA ATGACACAC ATCTCAGAA ACTCAGAAA TCATACTCTC AATGACAGG 300
TGTTCAGATG AATTACTA GATTCTCTT TGAAGGTCAG AATATGCTG ATATCATAC 360
TCCAAAAGAA CTGGGAATGG AGAAGAGCA TGTGATGAA GTTTATCAG ACAAACGGG 420
GGTCAATCA ACGATTAGA TATCTTTTAT ATTTTCTTTC TTTTCTCTCA ATCTTTTCTT 480
ATTTTATAAA ATAGTTCTTT TGTATGTGCG TTTTCAAAAC GGAATTTGAA ACTGCAACC 540
CATCTCTTGG AATCACTCTG TAAATTTGAT TCTATGCTCT ATATTTCAAT ATGTTTCTT 600
TTCAATGTGC TGAATTTTGG TGAATCAGCC TCAATGCGCT TCATATTACC CTCTCTCTTT 660
TAAATATTAC GTGTGACAG AGAGGTGACC TTTTTCAGGA CATTCGATTT TCAGGCTTGT 720
GCTCATTAAT AAGATCGACC AATCGAATG TTCAATATGA CTTTCCAAAT GCGCCTGATG 780
TTCTAGCATG TGAATTTCTT ACTCTGTGAC TGTGACTTTC AGTGGGAAT GGAATTTTCT 840
CAAGGAGCTG AACTGTGAAA AATGACCTT TCTTAACTT GAAGTACTT TTAAATTTTG 900
AGGCTCTGGA CCAAAAGAG AGGAATATCA GGTGAGATC AGATGACAG ATAGGTGAG 960
AGTAAATGAT AACTCCAAAG ATGCTCTCAC TGAAGAAAG GCATTTTATG ATTTTATTA 1020
AATCTGTGCA GAGATATCCA GAATGTCTT AATTTTCAAT AGCAATTAAT AAGATATAC 1080
ATGCAATAT GATATACCA GAACTGCTT CTTTGTGAT TTATTTGAT TTTTGTGCT 1140
GGAATATGGG TTTTAAATGG AATGTGTG TACCACTTTC ATTAAATTA ACATA 1196

60

(2) INFORMATION FOR SEQ ID NO: 223:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1791 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

5
10
15
20
25
30
35
40
45
50
55
60

TCACGAGAGT GECAGGAG AGCTTGACCA GCTCCGAG TGACGAGG GCGGCGCGCG
CGGTTGCGC TGCAGTTCGA AGCTTCCAGG TAGCGCGCGG CAGAGCGTCA CCGAGCGTCT
GCAGCTCTCG AGCCGAGTC CCCGACTC AGTGCAGTCA TCACTGCGGA AGTCAGGTCG
ACAGGCGGAGA ACCAGGAGCA ATTTCCTGCT CTAGCCAGT CCGCCAGGG GECAGCGCTG
GCCACACTCA TCCATCAGGT GCTCGAGGCG CTTGGTGTCT AGCTGTTTGG AGACTCTCTG
GACATGCCCA ATTCTTAGGA GCTGCGTAG AGTACCTTTG CCTCTACCTT CCGGCTGCTC
ACAGTCTTTC CTATTGCGAC ATACGCTGAC TACTTAGCTG AAGCCCGGAA TCTTCTGCTCA
CTACAGAGG CTCAGAGAA TAGGCTTCCA CACTCTTAG TTGTCACTCT GECTCTTAA
GTAAAGCTTA TCCATATGCG AGTCTCTG GAGTCTCTGC CCTGCTTAT GTCCGCGAGC
TGCAGAGCTT TGTATTTAG GCTCTGTATG CTGAGTGTCT TCTGTGCTCC CTGACAGC
GCAGCCAGG GCTCGAGGTT GACTACAGCA TCGGCGGGA CATCCAGGC CAGGAGCTCA
GTCCATTTGC CCGAGCTCTG CAGGATGCT GTGTGGGCTG TAGGTCCTG CTGTCAAGCA
TTGAGGAGCA GTTAGCGCTT GCGACCGAC ACAGAGGCA GAGCTGCGC CTGAGAGCAG
AGATTGAGAG TAGGCTTCCC AACCTTAAAA AACCATTTAA AGTTACGAGC GACAGCGAG
CGCAGCGCAC ATTCTAGGAC CTGTAGCGAC ACCTGACTGA GCTCAGGGA CCGAGCTCTG
GCAGCAGCA GCGCCAGCA GCAAGAGGC CTCAGAGGC AAGGCGGCTCC GAGGAGGCGC
CAGATTTTGG TCCAGTGA ATTGAGGGA CTGTGTTTC CTGCTTGGG AGTGTGGGTC
CGAGTGTGCT GCTTGTCTCT TAGGAGTCTCT CAGAGAGCT TCTGTGCGCC TGCGCAGCTG
ATTATCTTAG GTTCATGACC CTTCAGCTCC CTTACGCCA AACTATGATC AGAGCTTCTC
TAGGAGGAGG KAAATATGAG GTCATGTCTC TCTGCTTGT TTTCTGTTTT TTGTACTTCA
TGTGTTGCTT TGTGCGCGC TGCCATGCTC TCTGCTTGT TTTCTTTAGA GCTCAGCTC
TCTGCTCTTT CATTAATGT CATTTAGTAG GTGGTATGCC CTGATGCGGG TCGCTCTCTC
TGCAGCTTAA CCGCAGGCG TTTTCTTCTC CAGCCATCC CTGCTTGTCT GATCCCGAGT
TCTCTTATCC TACCCCTGAC CTATTAGCA GCTCTGAG AGCCATAGGG CCGCCAGCTT
TACTCAGAC CTGAGATTC TGGAGCGGAC TCTGCGATCC CAGGATCTAC TGCATCTTT

1560
1620
1680
1740
1791

CATCTTAGAA TCTCTGACA CTACATCAT TTCTTTTCT CTCTCTGCGC CTTCGCTCT
GCGATCTCTG CTGCTTCAAC CCGAGAGCTT AGCATGCGA GCGTTTCTT AACATGTTGA
GAGATGATTC TTCTTGGCC CTGCGCATCT CCGAGAGCTT GATGCGATC CTGCAAGGCT
TTATCTCTCT TTCTGAGGTT TGTGCGGAA GCGAGGCTTA TATGATTTCT ATTAAAGAAA
AAGAGCTATA TATGATATA TCTATATATA ATATGAGCGA GAAATTAATC T

(2) INFORMATION FOR SEQ ID NO: 224:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2517 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

25
30
35
40
45
50
55
60

ACATAGTGG ATCCAAAGAA TTGCGACAG CCGCAGGCA TTGTGAGCT TTCTCTGTG
TGTGCGCGC TCGAGCGAGC TGCATCTGTC CATCTCGGG TAGCGAAGTG GTTCTTTGTA
AAAGAGGCA TCCCATCAC GCTATAGGA AGTACAGAA CCGTGTCTG GACCGGAGC
CGATGAGTG CAGCTTCAAC ATGATGCGA ATGTTATCAC CTCAGAGCAG CCGATCTGC
TCCGCTTAG TGCAGCGCA TCATGAAA AGAGAGGCA GCTGCTTCC AGGTGAGCT
CTGCTCTCT CTCGACGCC CCGCGGAG TGGACCTGTA CACATCTCG AAGGAGCTCT
TCAGTCTCT AGGCGGCTCT ATGACAGGC AGCCAGAGA WTTCAAAATC AAGCTTTTGA
CAGGCGAGTT AGCGAGCGAG AGTGGCGGC AGAGAGGCTT GCTCTCTTTT CCGCAGGCA
AAGCTTATGA CCAATGGGCC ATCGGACTG AGACCGCTGA TTGTGGGAG GATTGCGAG
GATTAAGAGC TTCTCTACTG GATGCGAGCC GCGTTTCTCT GTTGTGTCTT GCGCTGTGCT
CTTCTCTCTA CATTAACTTT TCTTATGTA TTCTTCTTCA TCTCATGCGC AAGGTAGCT
TGTGTTTTTH AGTGTGTGCC TCCCGAGGC TACGCGGCA GCTGATTTCT TATCTGAAA
TGTACACTG AATCTCTCTG GTGGCTTTCT TGTGCGGCA TGGGATGAG CATTGCGGCT
GTCTGAGGA CCGTCTTTTT TCGAGGCGGC GAGGCGGCTC CTTCCTTTG TGTGTATTA
GCTTTTCAA CATTGAGGG GATGAGAGC CTTGCTGTG TACGCGGAG CAGGTGCGC
TGCAGGCTGT GCGCTCTCTC TGTCTTTCTA GTGAGGTGCT CTGGTGGGG AGGAGTCTC
AGGCTCTTGG TCTCTCTCCC AGTGGCTCA GCGCTACTA GTGCGAGGG CAGGATGAG
CTGACCGCTT GCGAGAGCTC TACTTAGCT CTTCGCTTGG AGTCCGCTGT CATTCTGCGC

5 CAGAGGAAAT TCTCCAGAGT TCACTTTTCC CTATTCTCTG AGTTCTCTCG AATGCCCTAC
CCCAGCTCTC TTTCCTCTCT GGTGTCTTTT GCTGGAGGG GCTGTCTCTG TGAAGCTCC
10 CCGTTCTCAC CTGGCTCCG ACTTACAC ACCCTGGTTT TGTGTAGCG CAGCTCTCT
TCTGTGTGG CCTTTGAAG GCTCAGCTC CCNTTGTCA GTCTTTGGGT TTGAGCTTA
TTTGAATGA AGAGTCTAGT TTGTCTCG CTCTCCATTT CTGGCTCAG TTGTCTACAG
15 GAGCTGTCT AGGATCTCT GAGGCTAT ATCCAGCTCC CACGAGGGG CACTGTCTGT
TCCACTTAT GTAGTGTACC CCCTCATCC ATGACCAGAG GATTATTTTC CTGCTTGGC
10 AGAGAGGGG GATCAGGGG AGCAGGGAG CTCTACCGAG CAGGTGTTT CCCAGCATTA
GGCCAGACA GTTGGAGCA AACTTCAGG CCGAGGCTGT CCTGTATGA CAGGCCAGT
20 GTTGTCTAG AGTGTCTCC TGTGTGTGG GAGTGTAGAG AATCCAGCT GCGAGGCTG
GAGCCAGTT GGGAGCAGG TTCTGGAGC TCTCCAAAT CAGTAGCAG TGGTGGAAA
GCCCATGCC GAGATACT AGAGCTCC AGATTTCT TGAAGTTC CAGTGTAAA
25 AAGCAGAGA CTGATTTTC CAGGGCTAG TCTGTCCGC AGGAGGACC CAGGATTTGA
ACCGAGTTC AGTGTCCAG CTCTAGCT GCGCAGCAG GGAAGTCCA AGGAGGGGC
CTGTGTCTG TCCACTTGA GTTCTTAAA GATGTCTCT TTTTATCTC CTAACTCTT
30 CAGTGGGTG CAGCTCTC GTTAGCTGT GAGAGCATT CCTCCAGAC TTATCCCTC
CTGGCCGAG AGAGCATCA GAGGCACTA GAGCTGTGT TTTCAGGTAG TGGAGCGGG
35 GGGTCACTG CTGGCACTGT GCTTAGGTA GGGATGTAA AATCTCTCC TGCATGCTT
TATCTCTCT CTGATCCAA AGCAGTATC TTCTGTCTGT CAGCAGGTT CATTGAGTCC
AGCTCAGCC AGCTGGCAT CTGAGCTGG TGTATAGGT GACCATCTG TACNTTGAAG
40 GAGCTGTCT GCTCTCTCA CTCTATAGC AGTCACTTG GAGAGCGGG AGGAGAGGT
GTTGGCTAG TCTGTCTCC TCTCCACTT CCTATGCTC TATGTACCC ATCTGTCTT
45 CCTGTGAGA AGGAGAGAA GGGGATTA GAGATGAG GTGATATGT ATTACTTATC
CATTCTGAA TAAACATTTG TTATCTTAA AAAAAAAA AAAAACTCG AGGGGGG 2517

(2) INFORMATION FOR SEQ ID NO: 225:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2424 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

5 TTGTATCTAA TCGAGATTC ATTCTATGA CAGAGTCTTT CAGAGCTTTC CAGATGATGT
ATCAGAGAG TACAGCTTTC CATGTACTG GAGATTTAGT AGAAGTCTG TCAATATTTT
10 TTTCGTCTTT GAGTCTTCA CCGCTTATC TTGAGAGAA AGATGTGAA CAGCATTAA
TCCATGCGA GAGCGAAT GAAATTTGCC ATAACTGTT AACTCTCTTT AATCTCTATA
20 GTCTCTCGA ACTTGAAGAT GCTGTATAG ATGTCTCAA GGAAGTCTTA CTTTTGAATC
CCATGATTT TTTCATACT CTGCTGCTT TTCTACACA CAGCATTTGT ACTTACCATC
15 AGATTAATAT ACCATGCTCT CTGAGCTTT ATTTCTCTG TGTAGAAAT ATCAGCTAA
TAGAGGAAA AGCATATTT CCGCTCTCC GCTGTAACT CATATCTCC CTCTTCCCA
CAGTGTGGA ACCAGTTAG GCGAAGATG AGTTTATGA TGTATCTG CTAGACTACT
20 TCTTTCTTA TCATAGTTC ATCCATCTAT TATGCCAGT TCCATCTAC TGTGAAGAT
TACTGTAAAC ATTAGTTAG CTGAGTCTCC TACTGTGCTA TGAAGTTTC CACTTCTATC
25 TTCACTGTT CCCAAACTT TGGAGTGGC TATGCCAGC TCAATCTGCT ATGTCAAAA
ACTGTCAA GCTTTGTGT GAGTCTCTG TTTTGTGGA ATATATTA TGTATCTTAA
TGTATTAAG AACTTTTAA AACACACA TTGTCTACAC GTTCTGACA CATTTCTTTC
30 TAAAGTCTA AGTCAAGTG TTTTCTGAG CAGCTCTCC CATTTCTATC AGCAGCTTAA
TTACAACTT GATAAGCCG TATCAGACC TACAGTCTGA TTCTCTCAC CAGTGTCAA
35 TTTCAGAGC AGTGTCTCT TTAAATGGG ACCTGAGGC ACTGCTTTTG CTCTGTGAG
TACAGCTC CAAAGATTA AACCGCTC TAAATCCAG TCTGCCAGAG CTTTTAAACA
1080 AATGAGAGC TTGTCTGCA CAGAGAACT CACTCCAGA GAGAGGCC AAGAGAAAGA
1140 AAGTAAAGA TGTATGAGA GCACTCCA TTAAAGGGG GCTGTTAGC AGTATGAG
1200 AGCAGCTGT AGCAGCTCC ATCATGTACA TGAAGAGCA AACAGGGAG GTCTGTAGCC
1260 CAGAGGAC TTCTGACAT GAGAGCAGAG ACTGCTCAT TATTGATCCA GGAAGTGAAC
1320 AAGTCTTCC TTCCCTCTGA AATAGTCTG TTAAAGATA CCAATGAAA GTTCTATCTT
1380 CATTCTGAA AGCATGTCTA AATATCAGT CAGAGCATCC AGAGAGCAG TCCAGCATG
1440 GTAGATATGA CAGTTTAAA GATTTTAAAG ACCTCCACTG TTCCAGGAT TCTACCTTAG
1500 CCGAGAGA ATCTGAGTTC CTTTCACTT CTATCTCTG AGTTCTCTCT GACTTAGCTG
1560 ACTGTAGAG CTGTATGCG CAGCTTTCG CTTCCAGGA CCGTGAAGTT GCTTTATCTC
1620 TCAATTTGCG CCATTTCCGA GCACTCTTAA GTCATATGCA GCAACATGAC ATTTTATATA
1680 CCGTGTAG GACATGAA TCTACATCC ATGTCTCAC AAGGATATC TGGCAAGGA
1740 AACCAAGTG CTCTGTGCA TTAGTGTAG CATGTCTACT TTTTAGTCCC TCAGCCGCA
1800

5
10
15
20
25
30
35
40
45
50
55
60

CCCCCATCT GTTGTATATA GTTTGTGCTTA TTGTGTTTTG TGTCTCACTT TGTCCAGTCC 1860
TCTCTCTCTG AATGCGAAGA TAGATTATATA GCGTTAATTC TTGCTCAGGC AGACTCCGAG 1920
ATGAAAAA CTTGCATCTT CAGTACTT CTTAAGGCG AATCAGATA TGGATATCTT 1980
TTATGTAAAT AAGAGTCCAC TTATGTGGCT TTCAATTAAT ATGCTGTCTT GCGAGAACAA 2040
GCGTTCCCTA GCGCTGTACA ATGTAAATTA AACTTACAGC ATTTTACTG TGTATGATAT 2100
GGTGTCTCTT GTCCAGCTTT TGTACTTTAT AGAGCGAGAT TGCCTCCGAT CCGTGTGCTT 2160
CTTATATACA AATTATAGTT TACTGTATTA CCGACACACC ACGAGAAAT TGAATCTGTA 2220
AAGATCTCTC TTATGCTGTG CCGTGGCAT ATATAAATCG TCCTTTATTT AACHAGATAC 2280
CTGTGAGGA AATTAAGCAC ACTTGATGTA AAATATATTT TTTTATTTTT ATGTACATGA 2340
CTGATTAAT CTAATCTCTT GCACTNAATT AAAGTATGTT TGTATGACTTA AAAAAAATA 2400
AAAAAAAA AAAAAAAAA AAAA 2424

25

(2) INFORMATION FOR SEQ ID NO: 226;

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

5
10
15
20
25
30
35
40
45
50
55
60

ATATAGGAGC GATTAATCTGT TTACATCTGT TTCTTCTGGA TCCACTGACA AGCGGATAC 60
TAGGTGACAA GAAACAAAG ATCTTAATCA AAGAGGTCT TACAGCAACC CAAGTCTCA 120
TCTTCCATA GTTAAGATCA CCGGCGCTTG AGTAAAGCTA CAGGCAACAC CACTTCCGG 180
TTTCTCTTGC GCGCTGTGCC AGATGCGCG ATCAAGCCAC GCGAGCTGTTT GTGTCTGAA 240
TCCCGGTGCTT GAAAGCTAAC GCGCGACCCC CAGATCGTGA GTTGTGGGTT CAGGCACTGA 300
AGGAGATA TCAATCTCTT ATCCGATATG TGGAGAACAA CAGATGCTCT GAGAACGAT 360
GCTTCCGACT GGAATCCAAC AAGGAGGAA CTCGGTGGTT TGGAAATCC TGTATATCC 420
ATGACTCTCT GAATATGAG TTGACATGCG AGTTTGACAT TCTATACAA TATCTTACTA 480
CTCCCCAGA AATTGCAATT CCGTAGCTGG ATGGAAAGAC AGCAAGATG TACGGGGGTT 540
CGAAATATG CTTGACGGAAT CATTTCAAC CTTTGTGGGC CAGGATATG CCAAAATTTG 600
GACTAGCTCA TCTCATGCTT CTGGGCTGCG GTCCATGGCT GGCATGTGAA ATCCCTGATC 660
TGATTCAGAA GGGGCTCTTC CACACAAAG AGAAATGCAA CCAATGAGAA ATCAAGGCCAC 720

5
10
15
20
25
30
35
40
45
50
55
60

TAGCGCGG CAGAGGAGAC TTGTATAGCC TAGATACTA TTTTCTCTGT CATCACTTT 780
AACTCATTA ACTGCTTCCC CCGACACGCT CACCTCTTAG TTATTACTAA GTATGCGAG 840
TAGGCAATTC TGGCGAGAA ACAAACACAC ACCAAAGAT ACTGCTACTT AGTTTCTAAG 900
GCTGCACAGG GAAGGGAAG ACTGGGCTTT GGCACATCTA GAGTAAATTT ATATCCGCC 960
CGAGGTGAG CACATCCGA TTCTGAGGCC AGGGGGTAA CTGAAAGTGA GTACATATAG 1020
TCTTCTGCT TTCTGAGAT AACCATCAA TAAAGCTGC TTCTCTCGG TAAAAAAG 1080

(2) INFORMATION FOR SEQ ID NO: 227;

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1336 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

5
10
15
20
25
30
35
40
45
50
55
60

TTGCAATCAC AATTAATCGG AGGCAAGGAG GGGCAATTC ATCTGCGGG TCGCTGCATG 60
GCTCCGACGC TCTCTGCGGT TTGAGCAATG CGTTACAGCT GCTTACGCTG AGCAACAGATG 120
TTATCTCTTA TCTCTGCGGT TGAATCTGTC AGCGGAGCAC TCGCACTATC AAAGATATCT 180
GTGAATATCAC CAGAAACAT CTTGTAGATC ATCGAGGCTA CTACTGCAAT GACTTGAACC 240
AAGTGAAGTT TCTTTTCCA TGTFTTCCCA TCTGGCAAT CTTCGCCAAA GCAATAAATAG 300
ATCTCAAGCG GTGCGTCTTT CTCTATCTGT CTTTCTGCTT GGGCAATGAG ATCTCTAAGG 360
AATGTTTCCA GACAAATAG CTTCGACCTTC TTTTGTCTCT CAATCAAGTT GCGAGCAACA 420
AGTATGCGCG CACATGCGCC AGACCAATAC AGCTTGCAGT GGCAGAGTCT GATGCGATTA 480
ATGCAATGAC CCGTGACCTC CAGGATCAAT CCTCTGTCCA TGAATTCGAG CAGCTTCTCA 540
GTGAACAGCT TCTGCTTCTC ATTGATATA TGTCTAGGAC CTGGAAATTT GACTTCTGTC 600
AGTCTGAGCG GACCAAGAG CTCTCTCTGG TCAAGCAATG GATCCAGGTC CCAATAGAG 660
AGTGGCACGC CTGAGGGTT CCTCAGGCTC ATGCTCTGCG CCAATCTCTT TCCGACGAT 720
CTGAACCTTG ATGTCAAGT CAATCATGCG GAGAGAGCTG ATCCAGATTT CTGAGAGCT 780
ATAGAGGCGC TGTATAGCTG CTTGCGGAC TTCCATCTCC AGGGTTTCAG TTTTGGGCCA 840
CACTGCTCC GCGCTGAGT TCCCAAGCT CCAATTCCCC ACACTGCTG GCGCCATGCG 900
AGAACATTT ATGTTACAGA AGGGAGGCT CTCTGTGATG GAAATATGTT GCTGCGAGCT 960
ATCGAGTCA TCTTCTCAT CTCTCTATC CACATCATTA TCTTCTCAT CCGAGGAGAC 1020
AGAGCCCTGT GATCTGCTT TAATGATCAA SCCTTGGGCG TGAAGGATGT CACAGACTTG 1080

ATATATCTTC ACTGGTTTCA TGGGACTTC CTTTGTGGCC ATCCATACAT CCAGGTTTGA 1140
TTTCTCTCTC TTATTTGAGAG CAGACGGGTC CTGGGCTTTC CATTTAGCTG GGTTCAGGGTC 1200
ATCCAGCCCT TCTCTGTACT TCTCTGTCTC TACAGCCGAG GCTTTAATAA TGGTATTTTC 1260
CTCTTTCTGT TGAAGGGCTAT GCGGGGTGGC ATGTTTCCAG GGAATCTGGA AGCGTTTGA 1320
GTCCCTGTGT AGCCAG 1336

15 (2) INFORMATION FOR SEQ ID NO: 228:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2043 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

25 TACAGCTGTC CTTTCTTTGT GTCTGGGGG ACCTGCTGGC GGCCTCTTTC TGGGAGCAT 60
GACCTCAGAC CCGACGCGCA CTCAGATTC AGACCCCTGC CTCGCCCGG CAAATCTCTT 120
CCGCTGCTCT TGCAGCTGTC ACTTTGACAA TGCCTACCCC CAGCAGATTC CCACTGGCCC 180
CTCACTGCCC CTTTCTGTGAG CTCTCTTCCA AGGACTCTGT GTCACTGCTT GCTGTGCAT 240
CAGAGCGCCA GGTTCGACGA GCGCCGSSGG AACGGTCTCT GCTCTTTCTT CCAATTACCT 300
CCAGTTGAGG TGTGAGACCC GTTGTGATTA AGGTCTGTAG CAGCGACGCT GCGCTCTGCC 360
CAGGCTTGGG TCTGTAGGAG CTGTTTCTTC TGCAGGAGG TTGAGGCGAG CAAATCTCTT 420
CTCTGCTCTC AGGCTCAGCT GCGCAGCTG GCGCGGATGC AGAGAGGCGAG GTGCGCTGTG 480
GCTTGAAGTG TCGGAGCTG GCTTTCTTAC CAGAAAGGCC TCAGCCCTTC TCTGTAAAGA 540
TCCGCCCTTC TGGCGAGGG GGAAGGGCTC CTTTAAAGGG TGTGCTTTTC CAGTGGCGAG 600
CAGTCTGGCC CTGCCCCCTA CTAAAGCTTC TGTCTGAGC ACTTTTCCCC AAGTCTCTGT 660
AACTTCTGTG AAGTGGGTTT GTGGCTGGCA GCGATCTCTT GCGAAGACTC TCTTGGCCCT 720
TTTAAATTTT ACTCATTTTG TATTAACCA GCAAGCTGCT GTTTACTTATG CCGCTATAGT 780
TTTTTCAATT TTCTTTTCCG TCTTTCTCTT TGAATCTGAG GTTCAATAT 6CTCTCTGCG 840
CCTGTGAGG GGAAGTGTCT CTTTCTGTGC CCACTGCGG GCTGTGTCCA GCAAGCGCTG 900
NCGCCAGTGG GGGGGCGCGG ATGGGGGGTT CTCTCTCTGG GAGGGGTGCA GGTGCCCTCC 960
CAGGGCTGGG AAGGTTCTCT CCGTAGCTGC CCACTGCCCC CCGCTGTGTA GAGTTGGGCT 1020
TCTTGTGTTT GAACTCTCTT GCAATTTGGA ACAGAGCTTT TCCAGCATTT GTTGTGTGTG 1080

TTTTACTCAC CTAACTCTTA GAAATATGAT GTTAGAGGT GCTTGGCGAG GCGGACAGA 1140
GTGTTTCTCT GCGCTGGAGA AGGCTCTGCT CAGGCTGTGAG AGTCCCTTTC TCGCCCAAGG 1200
ATACTGCGAC TTATAAAGGG AAGCTGAGCG CACAGTGTTC AGACGAATG GCGCCGAGAA 1260
GATGGGAGCT TCTGTCTGCG CTTTCTGTGT CTCTGTGACC TCAGCTAGCC TAGAGAGGAG 1320
GTGCATTCAG GGTAGATTTG CTTCTCAATC AAGTCTCTGG GCTTTGGGY GGAATACAGC 1380
CAGCTTTGGC GCTGTGSGGG AGACTCTCTC AGACAGGAA CCGCAGAGG AGACAGAGCC 1440
TGCACATCC TCCACAGCCA GCGCTGGGC CAGGGTGATT GAACTGAGAA TTGTGGCACA 1500
ACCAATTTGA TGTGTGCTGG AACGAGGCC CAGAAAGCTT GGCCTTCTTC CCAATGTGGA 1560
GCGCTGTCTT CAGGCTCTTT GTCCCTTTGA CTTAGTGA TTCCCAAGAG GTGCCACAG 1620
CTCTGAGCT TCATATCTA TATATTTAGA GAGTTGAGAA GTATATCAGA GATATTTTTC 1680
GAAAGCATTT GGTCTATGCA ATGTGAGTTT GGAATCTCTT TCAGAGTTTA ATGTTTTTAT 1740
TAGAGATTTT AAGAAATATA AAGGTCTACA ATATCTTTAG GTTTTTTTTT TTTCCTGTTT 1800
ACCGACAAA CTGACCAAT GGCATCTCTA TCAGAGTGA GGTCTCTCAT GTTCTCTCTT 1860
GTCTTTAGGG AGGTGATNAG CAGATGSSGG AAGGGTGTTT TTTTCTTTTG ACTGCCCTCC 1920
TTTCTAAACG AATGTTGCA CCACTGCTTG AGTGGGCTGT GTTGTCTCTT CTGTGCCAGC 1980
TTCTGTGTTA GAAATATACA TTGTTAGGGG AACTAGGSET AGTGTGAGCG TCTGTGTTTG 2040
GGG 2043

(2) INFORMATION FOR SEQ ID NO: 229:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

TAAGAAAG CCGGAGATC TGGGCTTCC TCTAGAGATC GATGGGCTAG AGGAGAGCT 60
GTCCAGTGT CCGAGAGAC TGGAGGCGGT GAACTCTAGA CTCACAGCC GCGAGCTGAG 120
CCAGAGGCC AGAGGTGCC TCGAGAGGGA GAAAGAGCC CTATGACAA AAGCTCTCAA 180
CTAGAGAG GAACTGAT TTCTTGGGCA AGAGAGCCG AGAGCATGCT TGTCTCTGT 240
GCGCATTTTT ATCTCTCTGA CCGCTCTCTA TGCCTACTG ACCATGTGAG CTTGCGACTT 300
CCCGAGACC AGCAGAGCT TCACTTGGC CCGTTGTGTA GATATCAGA GCGACTTCAA 360
GCTCATNAG GACCAAGGT CTGGGTTGTT CCGCTGCCAA CTTAGTGTTC AAGCATGGCT 420

TCCTGGGGCC CGAGGCTTGG CCTCGCTGGC CTCTGGGGGG GTTCCGGGTC TCACAGAGCA 480
CATCGTCTGG GTCCCTCCCT TCACCGAGCG GAGAGGCAAT AAGACACAA AGCTGGAAAT 540

5

(2) INFORMATION FOR SEQ ID NO: 230:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

AATTGTGAAA TATTAGATA TTGTACTAT TTGACCCAC TCMAATCTC CATTGGAAAA 60
TACTGTGCGA TACCACAGT ATTGTGAAA ATATACAGAT GCATATCAC AGCTGTGTCA 120
GACTCTAGTA CCAGTTGGCC ATTCAGGCA CAGCTAAAAA TTGAAAGCA AGATCTGAC 180
AAGCAAGAG CAAAGGTGG GGTGTGAAA GCTCTGAGCT GTACTAGCT GTAGAAATCT 240
ATGCACAGT GCCAGTGTG GTGTGCTAT CCAGGAAAA CTGAGAGAG CCCAGTCTTT 300
CACTCTGCTT TGACCTGAG CTCTGTGTA GCAGGAAGTG AAGGCTAAGC CAGATTTAG 360
CTCTGAAGC ATTCCACAC ATACACAAA ATCTGCGAAA GCATTAGCA ATCTCTTTA 420
CTCTAAGTG TTCTGACCC AGGAAACA 448

35

(2) INFORMATION FOR SEQ ID NO: 231:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 407 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

GTATCTGCC CCAACCAAT ATGTGTGCT GCTTTAGCC TGACTTCTCC AAGTGTAGC 60
CCCAAGAGCA GGCCTCTAGA CTAAGGAGG GCTGTGTGAC CCAGGTGTGG TGGGGCTGCA 120
TGAACTACC AGAGAGACAG AACTTCTGGA ACTGACCCCTG GGGGATCCAG TGGATCTGCC 180
TATGCTCTGG TCCACCCGAG ACCTGTGGA TTTTCTCTAT GAGGTGAC TTGCTCTTCT 240
GCAGTATATG CTCAGCTTTC ATATGACTC CCACAGCAC CAGCAATACA GTTACCTACC 300
TGTGGCTTGG GATCTGACC AGCATGCTG GAGAGGGGAG CAGCTGGGCA TGTACCTTAA 360
ATCTGTATAC CAGGGAAGA CTCGAGAGT GAGACAGAT AGGGACT 407

60

(2) INFORMATION FOR SEQ ID NO: 232:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

GTATTGATTT TCAGGCTGCT AATGGGCTC ATTAGCAAT CATTCCTTCA TGTAGCAATT 60
AAAAAAAAA CTGANTACA TTCTTTCCAG GATPACTAT AAGGAGACA TGTAGGCTT 120
ATTAATACAT CAGCACTGCA GCACACGTTT AAGGTTGCA GCGAGAGCA TCACACATA 180
GAGACACTG TATGTCGCTC TCTCTCAG AGCCAGACA TTGATCAGTT TTGTGTTTG 240
GTTTATATTT TTCTGTTTAA AAAATTTGCA AAGTTTGTTT TTAGCTAGAT GATATTTTAA 300
TAGCTCGGAG TCTTTTGAAA CTATPAGAT GTCACTACTT AAGACATA CTTTATGTTT 360
TTTTTTGTTT TTTTATAC TCAGTATAA TCAGAGAG TTAGCCAGC ATCTAGCAAT 420
TAGATCTCTC TTTTATATG TCTCTCAG ATATGATGT TCCCATACA GCACAAAAAC 480
AGCAACAAA ACAATTTGTA AATATCACTT GATAGACTGT AAGCACTGC TTAACTTTGT 540
GTGCCAATA TTTAGTGTGT ATATATAT ATATATAC AAGACACAC ATATATATTC 600
AACAAATAAA GCAAAATATA ACATGCAATT CAGATTTTGT CTTTCCCTGT TACGATTTTA 660
ATACAGAAC TGTATGACA GTTAGGCTCA TCTTAGCATA TTTAAATTC AATTATAT 720
AAACACAGAT AACACACA ACAAAGCTGT CTATTTGAT GAGTCAATC TTCTATAT 780
AATAACTTGG CTTCGTTAT CCAATCAATG CACACTTATA CTGTTATCTG 830

45

(2) INFORMATION FOR SEQ ID NO: 233:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

CCAGAGAAA GACCAATCTA GATATGAAA CTCATATCAC TTCTATAT TCAACTTCTT 60
AGCAAAATG AACTTGGCCC TAACTTAGG GATPAGACA TTTTCTTAT GTAGCAATG 120
CTACGAAAC AAAAGAGTG AAGAGGCC TTTTATTTATA CTTATATGAC ATATATGAC 180

180

5 TTTTGTGACA AGAATGCGAG AATATAGCCTT CATTTCTTACC CTGCAAAATA ATCCAGATCT 240
GCCTTCTAAA ATGATATCGAG TTTCATAGCT GAACATGCA AATATTTAGC TGTGACTGAC 300
TCTGATATG GAGGAGGAG TACTTCTGTT TACAGAAAC GATATTTGTA TATATGTCAG 360
GCTGTGATTT CTGACTATCA GCATCTGCT GCAATGACAC TTTCCTCCTT CATGACTGCT 420
GGAATTTTGA TACTTTTAAA GGTATTTCTT CTATGAGCAC AGCTCTCTCT AGTGAACCTT 480
AATTTGACAA AGGTGTGCTT ATGCTTTCTT AACCTGATTT GTATTAACTT TCGAGAGCC 540
TACATTTCTT CATTAGGGTT RTGATGCTCA GTATCTTTCC AAGTGGCAGG CAGRGCTTNC 600
CTTTTCTGAT CAACATATCC AATTTTTGTG TTTCACACT ATGAGAGTCT ACTTCTGAG 660
TCCCAATTTA AAATGCGAGA ACTGCTTTAT CCGAATGTC TCGAAATATC TGTTCATACC 720
AGGTTTCTTA TACTATATTA GCGATTTTGT CTTTGTGTTG TTATGCTTAG GCAATGGCGA 780
GCAATGTGGA TTAGCTTAG GCTTAATATC AGATGCAATCT CTTGTAACTT GACCACTGTC 840
TCAATATCAA GAGCTGCGAG AGCAATTTTC CAGAGCTGTG ATTGCCGAGA ACACATATTC 900
CCACAGTTTC TAAATTTGAG CAATCTTAAA AG 932

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2786 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 234:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

40 TTAGCAGGCT GAGCTGTAAA AAGACGACAC ATCTGTCATC CCGTCTTCTT TTATTCGCC 60
CTGGATTTC AAGAGGAGG ATATATGGGG ACGACTGCC CTTCTTTTGA TCCGAGCATC 120
TCAGTCCCC TCCCAACCTT CCAATGCTT CTCAATGGTG CTGACTTCTT TCGAGCGGG 180
CTCCCAATAG CGAGGGGCTT GCGCTCTACA GTCTCTTTTGA CTGTATAGACA GCGCTCTGTA 240
TCATGAGAC GATGAGAAA GTCCAGGCT AATGCGGAA ATTTGCACTT TCGAATGCTG 300
TGTTTTGTG TTGTGGAGC TGAATCTCTT TATTTATTTA CAGGAGCTT GATTTTATTT 360
TTTTCAGTC TTGTGTTGTA TATTTGTGG GCGTGGGAGA GAGGATTTAG ATTTATTTGA 420
CATGGATGCC CTTCATATAC AGTACTTTTG AAGCGAGAC ATAGGGTTGA AGAGAGCAA 480
CGAGCTCTGG AATGATNAG TCTCAATGCG CTTTATNAGA AAGCTGTGCC TCGAGCTTAA 540
CAAAATCACT ACGATAGCTT AGTGGTTTTT TCGAAGCCTT TTATGGGAG AATGTTAGCT 600

TCATGTATAC TAGATATGTC TTTCAGATTT TTACAGTCTT GAACTTATAG AATTTTGA 660
GGTGGAGAG GGTGTGTGAG AATCTAAATT ACGATATAGT GATTTGTTCT TGTGAATTTG 720
TTTCTTTTCC TTTTATTTTG TCCCTACCAT TTCTTTACAT TTCCCTTTGG GCCCATCTCT 780
GGCTCTTCC TTTTGTGTTT TGTCTTTGCT TTATCGATGTC ATTCGAGTCT CCGTTATGTC 840
AAGGACACTG CTGTATGTA AGGAGAAAG TCTATGAGTC CTAAATTTT AAGTCMAAGA 900
AAATGCTCTT GTTTCCCTCT TATGTAACACT TCTGAGAGAG AAAAATCTTA ATAGCCAAAG 960
TTATATATCC TATATATATA TTGCTTTTGG TTTCACCTAA AATTCCTGGC ATCGAATTT 1020
CTTTGGGATA GAGGTGTGTT TGGGAAATAG ATGCTTTATT CCGTTTCACT GGAAGAGAAA 1080
GGTAGTGTTT TTGTACAGG TCGTACGGC AGAAGGCCCA ATCTCTATTT TCGCTCATCT 1140
TCAGTAAAG AGTAAATCTT ATCTGTGTG CCGTCAAGGC TAGAATGAA GCGTTTACCT 1200
ATTCAATGTT TATGTGCGA AATGCAATGAT GCGCTCTGGA AGAATGAGG TTTTGTCTGA 1260
AAAAAATA AGACAGATTT GTGTTCACA AACTGCGCTT ATCAATTTT TCGAAGATTT 1320
CTTTTTCGCC AAAAGAGGA GTATCAAAAT GTCAATTTCTG AAGAGCGCTT ACTTTATACC 1380
AACTAGTCT AGCAATTTGG ATCGAGGGA ACGAGAGTG AGACACTTAC ATCAACCAT 1440
CTCAATGCG CTAATGTTTC TTTCAGAGT GTTCAGATTT TCCCATTTCT CCAATATATG 1500
GGATATGAA ATGCAATATA GATGAGAGG ATGTAGATA TCGTTTCTG CCAAGATGCT 1560
TTGAGTGA CTTTGTATTA TTGACTAGAT TTGAAATATC AAGATGATTT AGATGAATCT 1620
ACAAAAAAT TGTCTGTC TCAAGTCCCT TTTCACCTTT TTGACTACTT AGCATCTATA 1680
TTCCACACTT AGCTTTTGTG TCGACTTAT CTTTGTGTC CATTAAATTC ATTTCAGTG 1740
GTTAGTCTC AATATTTTGA GCGACTACA CAAAGAGCAA CTGCATTTT AAAAATCTTT 1800
CTGAGATGG AGAATATGTA TTCTGCTTTC CTATAGCGCT CTCCCAACA AAAAACAAT 1860
AGTAAATCTT ACTATATGA AACTGCTGTT ACTTTTCTTT TTCTTTTAGG GGTCAAGCAC 1920
CCTCTTTATA GCTACATTT GCTACATAA AATATATGA GCAATTTGCA ATACTAAAT 1980
ATTTTATTA GACTTTATTT TTTCCTTTT CATTAAGGGA TCGTGTATAG TAGAGTTGCT 2040
GTATATTAAC TATCTGAGC GTTTCCCTGC TTTCCTTTCT CCGTCAATAG CCGTATGTC 2100
CTTTCAGGA GCCTTTTAA TCTTAAGATT CTACATTTCA TCGTCTTATG CAATATCTCT 2160
TACCTTTTAA ATACTGTC CCACTGATA TTTCATCTTT GATTTGATGG TTCTAAATTC 2220
TGAATCTGTA GTTGAGATAC AGCTATTTAA TATTTCTGGG AGATGTGAT CCGCTCTCTT 2280
KGTGCTGGC CAAGTTGTT TTGCTAACT GAGACTCTCTT GATATGCTTC AGAGATTTTA 2340
GGCAAACTT GCGCATGGCC GTGGAGTAC TGGGAGTAAA ATAAAAATAT CAGGTATAG 2400

ACTAGACATCC ACATAGAGACA CTTCAGCTC CTCTTACTCT GTTTGGGGAA AAGATATAT 2460
GAGTGTACTA CCATCTACAC TAAGTATATT ATAGTCTGCT TGTTCGAAT ACATTTTTT 2520
5 TCCTGTTTG TGTTTTCCC ACTTTCCAT GTACTCAGA AATTTGACA ATGTATATG 2580
ATCATTTTAA ATATTTTAT TTCTTAAGG CCTTTTTCG CTGTTGTAT GTCCAGGACC 2640
CTTCTCTTT CATGGAGAG ACAGTATATT ACCTGATAT AGTTTGAAA GTTATGTAA 2700
10 AAGAAATTA TATTAAGAG GATCTTTCG TTTTCAATC TTGTGTTTCT CTATTTCAG 2760
GTAAAGCAT TTAATAATTA ATATGT 2786

15

(2) INFORMATION FOR SEQ ID NO: 235:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

GGGTGCAGGA ATTCGGCAG AGAATATGTT TGAATTTCT TCTATTTTAA AGATCTTCT 60
30 CTCTCTTGA TGTTGAAAC TTACTTACT GAGATGCT TTCAACATC TGTGTCTTT 120
TACCTGATA ATCAAGCTA TGCATCTATT CAAGTGTATG ATCTGTGGA TAGTTTTAT 180
35 GAGTCACAA ACCAACTACT AGATTAAG AGATGATCA AACCTGGAC CCTCGAAA 240
GGATTTCTT TAGTGACTGT TCMAAGAA GMAAGGAACT TTTTATACA ACAAAGAGA 300
TTCTTTTAA ATATGAACC TGAATTCAG CCTTCAGATA CAGGTACTT GCGCTCTTC 360
40 TTTTCATCC ATCTCTTTG CACTCTCAG TGGAAATATT TTTAAGTGT TTAATATCT 420
AAGTCTTGT GAACCTAAC AAGATATCC CTTCCTAA 458

45

(2) INFORMATION FOR SEQ ID NO: 236:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 591 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

AGCATGAGA GGAATATATC TCTTGATTC CTTCACAGA AATCTCTTC TATTTCTTAA 60
AAGCTCTGT TCTTTCTAG GATTCATG TGTGATTC TCTTAACAT CAGGTACAA 120

TTAAGTCTT AGATTTGTA TGAAGGTTA ACTCAGTCA AATTTACTT GATCTCTCTG 180
AATATATCT GCACTGACA ATGAGAGAG AACAGAAA TGTATGTGA TGTCTCTCC 240
5 CAATCTCAT ATCGTTTTC GATTTGTTT GATPATTTT TCTTTTTTC TTATCTCTCC 300
TTATAGCC TTTCGACAT TCGATATCC CAGCCACTC TCCACATCA ATGTAACTCC 360
ATGCAATG CTCTCTCTG TGTGTATC TAATTTTTT GATAGGAAA CAATATCTTT 420
10 TGAATAAA TAATATACA AACATAAA GTTATATG CCAAGTTTA CCTTGGAAG 480
TTTTTCTAA ATCGGAGAG AGATTAATCT TTTTAAAG TACATATAT GAATATAT 540
15 GTAACTGTA ATATTTGTA AAGTCTGAT TTCCAACT CTTTGGGAT T 591

(2) INFORMATION FOR SEQ ID NO: 237:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1286 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

TCCTTTTAA GTACAGAG GAGAACTGG AACTCAGAG AAGAACTG CCTTCCATC 60
TACAAAGCT GAGTTACTT CTCTCTCTC TTTGTGAG ACTGGCTTC CAGCCAGAG 120
35 GAGTTACTT GGGCAATG ATGTATATG TCAGACTATA ACTATCAGCC GAGTAAAGG 180
CAGCCAGAG GCAATGAGA AGCAATCAT ACAGTCTTT TGTAAAGAT CTCTACTTA 240
AATAGAAC AATTTTACA AACCACTCC GTTTTCTCT CAGGAACTC CTCCCACTCA 300
40 CCTTCACT CTTCATTTT TTCCACTCC TCGACTCTC AGACTCTCT CACTCTGAT 360
TCCAGAGG GGTTTTCTC CTCACAGG CGCTCCACT CATCTCTTA TACCACAT 420
45 AGAATGGA CATTCCTCTG GTTATGATG TSGTCTGCA CTGCTTTTC CATATGCGAA 480
TCCGATGAG AACATACAG ATACAGGAA TATCCAGAA GAGTTATCA CCGTCAAGA 540
GCAATGAG AAGAGAGA AGCAGATAGA GAAGAGAC AAGGAGAAA AGGAAACCC 600
50 AGCATATAT CTTCGAG TAAATATAGA CTTGGCATG AAGTGTAGA AGCATATAT 660
CAGCAGAC ACATACAGA AATATCTTAA AGAGCAAG AGGAAAGA ACCGGGAGT 720
75 GAGCTTCCC CTGAACAGA GAGCAGCAA CCACTACCTG CAGATAGCC ATGCTTTTGG 780
CCTTTTCTT ATATATAC CAGATATAGA TACTATAT CTGTTTAT TTCTGATAAA 840
TGTTAAGA ATTTACTTAA ATCTCTCTC TGTGTTTAT TATGAAGAT TAACTTTTTT 900
60 TCCAAATTA AAGATGAT TTTTCAAT TTTTGTCTT GTACTATTC 960

AAAAAAAAA AGACAGCAT GACTTATAT CCAGAAAGG AATCGAATG AGTCACTTA 1020
5 CAGGGAATCT AAGAGACTGT GTTACTGTG TACATACCA GATTATCTGA GAAAGATCA 1080
AGGTTTCAC TTGGGACCA GTTTTATCT TAATCAACA CCACTCTCTT AAGAGCTGC 1140
ATCAGAAAG GCACCAAGG GGGGCTCTT ARGGCTTTGA GATTAAAC TACTCTTAT 1200
CCATTACTG TGTGAACTT CTGGCTTGG TATTTTATG GGGATCTT ACTTTTAT 1260
GGTTTTCAC ACTTTTGG TTGGC 1286

(2) INFORMATION FOR SEQ ID NO: 238:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 734 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

ATGCGACGC AGAAGACCA GCGAAGAT GCGAGGCGG AAGGCTGAG CGGACGAC 60
CTGTGCGCA AGCTATTC CTCCGTGCA GCGGCGAGT GCTTGAGAG GCGCGCGCG 120
ACCATCGGC CTGCGACAC CTTCGTGAC CAGCAGGCT TCTACGCGC CCGCAACCG 180
GGAGACTGT CCGACGCT CTACGACAC GTGAGTACT ACGAGGAA CTATGTCTC 240
GTGTTCCTG GCGTCATCT GTACTGTGT GTACGTGCG CTATGTCTG GTTGCTCTG 300
GCTGTCTTT TCGGCGCTG TTACATCT CTATCTGCG ACTTGAGT CCAAGCTTGT 360
CCTCTTGGC CAAAGTGA GCGACGCA TCATATGCT TGGCTGAGG CATCTGCTC 420
CCCTTCTCT GCTTGGCTG TCGGCGCTG GCGCTCTCT GGTCTCTGG AGCGACCTG 480
GTGTCTATG GCTCCACGC TCGCTTCCAC CAGATTGAG CTGTGAGCG GAGGAGCTG 540
45 CAGATGAC CCCTGTGAG TGTCTCTGG GACTCTGCG CTTCGCGCG CAGCTGCGCC 600
ACCTCTGCG ATGCTGTGC TCGAGGCT CTCTCTGCG CCGACGCG CTTCCATCA 660
50 CAGCGCGCG GAGGATCC GCGTTGAA ATTAAGCTT TATGCTGTG ATTCAAAAA 720
AAAAAAAAA AAA 734

(2) INFORMATION FOR SEQ ID NO: 239:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 809 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

5 CGGGCTTTC AGGTAACCG GCTGTTTACA CGAGCTCTAC CCTCAGAC GCAATCATG 60
CAGCGCAAA GACACAGAG AAGATTCGG AGGCGAAGG GCTGAGCGC AGCACTCTC 120
10 TCGCAAGCT GATTCTCTC GTTCAAGCC GCGATGCGT GAGCGCGCG CCGCGACCA 180
TTGCGCGTG GAGCACTTC GTGACAGC AGGCTTCTC ACGGCGCGC AACCTGGAG 240
15 AGCTGTCCA GCGCTCTGA CCGAGTGG AGTACTACA GACCACTAT GTTTCGTGT 300
TTCTGGGCT CATCTGTAC TGTGTGTA GATCCCTTAT GTTCTGCTG GCTCTGCTG 360
TTTCTTTCG CCGCTTTAC ATTCTTATC TCGCACTCT GAGTCTCAG CTGTGCTCT 420
20 TTGCGCAGA GGTGAGCCA GCGATCATG ATGCTCTGC TCGAGGCAT TCTTTCCTC 480
TCTCTGCTG GCTGTGCGG GCTTGGCGG TCTTCTGCT GTTGGAGCC ACCCTGTGG 540
TCACTGCTC CCAAGCTGCC TTTCACAGA TTGAGGCTGT GAGCGGGAG GAGCTGACA 600
25 TCGAACCGT GTGAGTGTG TTCTGGAGC TCGCGCGCTC CCGCGCGAG TCGCGCACCC 660
CTGCGATGC CTGTCTGCA GCGCTCTGT GCTTGGCGC ACAGCGCTCT CCTATCAAA 720
30 GCGCGGAG GATTCGCGC TTTCAGATA AGCTGTAT GGTGTCTAT CAGGAAAAA 780
AAAAAAAAA AAAAAAAAAA 809

(2) INFORMATION FOR SEQ ID NO: 240:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

TCGACGAG CTTCCGCGA CATGCGGCT GCGTGTGTC AGCGCGCGG CTGACGACA 60
GCAATGAG CCGGCTCTA CCGCGGTA GGGTGGCT CCGCTGGA TTGTGCTTC 120
10 TTGAGCGCG TCTGTCTCC GACATCATG GTATTCGCA GTTCCCTTC GCGCTGTGT 180
TTTACTGAA CCGGAGGCG TCGAGGCGC CCGTTCCTT CTCAGTGTG CTCTGTGCTT 240
55 CAGGCGAGC TCCCGCTCT CCGCGCAT TCTCTGCTT GTTTCGCTC CATCTCTCT 300
TCTCAGCT CTTCCTCTG CAGCGGATG AATCGACCA CCGGCTAGT CTTGGCACCC 360
CGGCTTTC AGGTCTCAG GCGACCGA AGTTCGAGA GCGCGAGTC CCGAACCGG 420

480	AGGGGCTGCA GGGCACTGAC GGGCTGTGTCT CCGGGCTGTTT GGGTCTGTATA GTGGGGGGTGG
540	TGTGCTTACAT CAACTCTCTGG TACTACTATGG ACCGGCTTGC ACCGGCTGGG GTTCTGTCCCG
600	ACATGCGAGCA GTTCTGTGAC ATGCGGGAGCA GTAGAGCTGTGG GCTGTATCGAG ACCGTGTGTCA
660	TCTCCAGTTA CATGGTGTGG GCACTGTGTG TTGGGTACCT GGGTGTACAG TACATTCGGA
720	AGTATTTCAAT GTGCGGGGGG ATTTGGCTTTCT GGTTCCTGTGT GATCACTGGGG TCAATCTTTCA
780	TCCCGGAGCA GCAATTTCTGG CTGGTCTCTC TGACCCCGGG CTTGGTGGGG GTGCGGGAGG
840	CGAGTTATTC CAGCATTCGG CCGACTGTCTA TTGCGAGCTT CTTTGTGGGC GATCAGCGGA
900	CGGATGCTC AGCATCTCTT ACTTTGGCAT TTGGGTTGGG AGTGGTCTGG GGTATCAATTC
960	AGGCTCCAAA GTGAGGATA TGCGTGGAGA CTGGCACTGG GCTCTGAGGG TGACACCGGG
1020	TCTAGAGTGG GTGCGGTTTC TGCTGTCTTTT CTTGGTATGG CCGGAGCGGC CAGCGGGAGC
1080	CGTGGAGGC CACTCAGATT TGCGACCGCT GAAACCCAGC TGCTGGTGGG CAGATCTGAG
1140	GGCTCTGCGA AGAATTTCTA GTTTGTCTCT GTTTTCCCTG GGTCTCACTG CTGGTGGCTTT
1200	TGTCAACGGC TGCCGTGGCT TGTGGGGTCC GGCATTCCTG CTGGCTTCCC GGTGTGGCTT
1260	TGGGAGAGC CGACCTGTCC TTGCCCGAGA GTCTGTCTCT TGCTCTGACA GTCTATCAATTT
1320	TGCACTCATC ACTGGCTGTA CCGGAGTCTCT GGGTGTGGGC CTGGTGGTGG AGATACCGCG
1380	CGGGCTTCGC CACTCTCAGC CCGGGCTGTA TGCCCTGTGTC TGTCGCATGG GCTCTGTGGG
1440	CTCTCAACC TTCTCTCTCC TGTCCTTTGG CTGGCGCCCT GTTAGGCTGG TGCGCACTTTA
1500	TATTTTATTC TTCAATTGAG AGACCTCTCT GTCCATGAGC TGGGGCATTC TGCGCGAGAT
1560	TCTGCTGTAC GTGGTGTGAT CTACCGGAGG CTCCACCGCC GAGCGCTTCC AGATGTGGCT
1620	GTCCCACTGG CTGGTGTGAG CTGGGAGGCC CTAACTCAAT GGGCTGTACT CTGACCGGCT
1680	GGCGCGGAG TGGGCGCCCTCT CTTTCTTGTG CGAGTTTCGG GCTCTGGAGT TCTGCTCAAT
1740	GCTCTGGCGG TTTTGTGTGG CACTGTGGGG CGCACTTTCC TGGGGACGCG CACTTTCAATTT
1800	GAGGCGGACC GCGCGCGCGC AGAGTGTGAC GTGAGGGGAC TGCTGCAAGA AGCAAGGTCC
1860	ACAGAGGAGC GATTTGTGTG GCGCCAGGGG GGGCGCTTCA CCGCGGTGCC CGTGGCGAGT
1920	GTGCTCATCT GAGAGCTGCG GGTCTCACTTA CTTGTGCATC TGCCACAGCT GGGCGCTGGGC
1980	CGACCCGAGG AAGGGCTTGG GCTTAAAGCC TTGGCTGTGG CCAAGTTTCA GAGGGAGCTT
2040	GCGCGCTGTG CGAGTCCCA GACACTACAT GGGTAGTCTA GGGGAGGAGG TGCGGGTCCA
2100	GGAAGGGGAT CCGCTCTCAC AAGGGGAGCC CGAAGGGCTC GGTGTCAATTT GTAAAGGAT
2160	AAGATTTGTA GCGAGGCC AGGTGGCTGG TCTGTCTTTT CTCTGGTGGG CCGCTGTACT
2201	TGCACCCCTT CTTACCCCA GGGGCTCTGA AGAGTGTGGG T

(2) INFORMATION FOR SEQ ID NO: 241:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1661 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

15	60	QTCCTCTCCG	AGATCGAGCA	GTCTCTTAC	ATCGCGGACA	GTACACTCTG	GCCTATCCAG
	120	ACGCTGTCTCA	TCTCTCAATT	CATCGCTCTG	GCACCTCTGT	TTCCTACT	GCCTTACAGG
	180	TACATTCGGA	AGTATCTCAT	GTCCGCGGCG	ATTGCTCTCT	GCCTCCCTGT	GCACACTCGG
20	240	TCTATCTTCA	TCCCTCGGCA	GCATTCTCG	CTCTCTCTCT	TGACCCGCGG	CGCTGTGCGG
	300	GTCCGCGGAG	CGAGTTATTC	CACCATTCGG	CCCACTATCT	TTCGCGACT	CTTTCTGCGC
	360	GCACAGCGGA	SCCGATCTCT	AGCATCTCT	ACTTTGCGAT	TCCCTTGGCG	AGTCTGTCTG
25	420	GCCTACATTC	AGGCTCTCAA	GTGAGGATTA	TGCTCTGAGA	CTCGACTCG	GCCTCTAGGG
	480	TGACACCGGG	TCTTAGGATG	GTGCGCTCTT	TGCTCTCTTT	CTCTGTATAT	CGGAGACCGC
	540	CAGAGCGGAG	CGTGGAGCG	CACCTCAATT	TGCGACCCCT	GAACCCACG	TGCTGTGTGG
30	600	CAGATTTGAG	GCCTCTCGCA	AGAAATCTTA	GTTTCTCTCT	GTCTCTCTCT	GCCTTACATC
	660	CTCTGCGCTT	TGTCACCGGC	TGCTCTGCTC	TGTGGGCTCC	GCATTTATCT	CTGCTCTCCC
35	720	GCCTGTCTCT	TGGGAGAGCC	CGACCTTCGC	TTCCTCGAGA	CTCTCTCTCT	TGCTCTTACA
	780	GTCTCATCTT	TGGACTATTC	AGCTGCCTTA	CGGAGTCTCT	GCCTTGTGGC	CTGGCTGTGG
40	840	AGATCAGCGG	CGGTTTCGCG	CAGTCCAGAC	CCGCGGCTGA	TGCTCTTGCT	TGTCGCACTG
	900	GCCTCTCTGG	CTCTCTGACC	TTCCTCTTTC	TCTCCCTTTC	CTCGCCCTCT	GCATGACATG
	960	TGGCGCATTA	TATTTTCTAT	TTCATTGAGG	AGACCTCTCT	GTCTATTAC	TGGCGCATTC
45	1020	TGGCGCGACT	TCTCTCTTAC	GTGCTGATCC	CTACCCGACG	CTCGACCGCC	GAGCGCTTTC
	1080	AGATCTCTCT	GTCCGCACTG	CTGGGTATAT	CTGGGAGCCC	CTACCTCATT	GCCTCTGATCT
50	1140	CTACACCTCT	GCGCCGAGAC	TGCGCCCTCT	CTCTTCTTTC	CGATTTTCGG	GCCTCTCATG
	1200	TCTCTCTCTAT	GCCTCTCGCG	TTTCTTTGGG	CAGTCTGGGG	CGCATTTTTC	TGGCGCACCG
	1260	CATCTCTCAT	GAGCGCGACC	AGAGCTGCAC	GTGAGCGGCC	TGCTCTGAGA	
55	1320	AGGAGGGTCC	ACAGAGCACC	GCATTTGTGT	GCCTCGAGGG	GCACGCTTCA	CCCGCTGTCC
	1380	CTGTGCGCAT	GTCTCTCATCT	GAGAGGCTTC	GCCTGACTTA	CTGCTCATCT	TGCGCACTCT
60	1440	KGCGCTCGGG	CGACCTTCAG	AGGCGGCTCT	GCCTTACCCC	TTCGCTGTGC	CGAGCTTCTA

CAATTC

1146

5

GAGGACCTT GGGCCCTGTG CAGCTGCCA GACACTACTT GGTAGCTCA GGGAGGAGG 1500
TGGGGTCCA CGAGGGGAT CCGCTTCAC AGGGGACCC CGAAGGCTC GGTCTATTT 1560
GTACGGAT AAAAAATTTA GCGAGCCCG AGTGGCTGC TCTCTCTTT CTCTGGGTG 1620
CTCTGACTT TCCAGCCCTT CTTCACCCA GGGCTCTCA A 1661

10

(2) INFORMATION FOR SEQ ID NO: 242:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1146 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

NGACAGAAA CGAGAGATG AGACTGTTT CATTCATTTT TCTTAGGCC ATCTGTGCT 60
CACTTTTCCC CCGTCCATCA TACCTCTCC TCTCTGGAG CTCTGCGGC TTGCTGTAA 120
TGTGTGACT TACCTGGATA TTTCAGTGG AGCATGAAG CGGAGCTCA CCGTAGCGG 180
TGGGAGCAT GGGAGAGGA AAGAGGAGA GATGCCAGG AGAGGGTCC AGGAGAAC 240
AGAGAGTTG GGTACAGGA AAGATGAG GAGAGATGG GTTCAGGCC CTGAGGGCG 300
GTTTACGAG CAGCTGGGC CTTCGCGGC CTTTGGCAT CAACTTGGG GACAGGTTAC 360
CAGGCTGCTG GTGTGTATCA TACGATTTT TCACATAGT TTATGATCA TCTTTGTAG 420
GTTTAGAGC CCGGTGTGT GACCATGACA TCGAACCCCT ATATATTAAG ATATATAT 480
ATATATATAT ATGTAATTA TACGACTGAG GCGCCCTGCT CCGTCTGGA CGAAGCAMA 540
CTAGAGCTTT TGTATTGCT ATATGTTTC GTTTGTATF TTGTTTGTAT TTGTGCTTG 600
TCTTATGTG TGTATGACA ATGCGAGTC GGAATGCTCT GTATTACAG ATAGTGTTT 660
TATTTCTCA ATGTGCTAGT TATGTCTAG CTGAGACCT CCGTCTAGC TATTTTTAG 720
AGGTTCACCA ATTTGTTTC TTGATTTTA CTGTACTT TTTTGTACA ATCAATCTCT 780
TTCTCTCTTT CTCTCTCCC CAGTCTGAC CATTGCGTC TCGATCCC CTCTCGCGCC 840
TCCGCTGCTC CCGTGGCTC CCGTCTCAT TTCTGTCCAC TCGATCTCT CTGCTCTCT 900
CGTGGCTCT GGTGCGCCCT CCGGAGCCA GTTSCCGAG TTGTGCTTCC GGTCTCTTAT 960
CTGTCTTAGT TCGGAGCAG TTTCATGCA AGTTGTCCAG TCTGTGTTCC AGCTTTTCCC 1020
ATCTGCTTC GTTTGTGTA GATGAGCGG TTCTTTGTA ATTTCAATCT TTCTGACAG 1080
ATTTAAAAA AAAAAAGGA AAAAAAAA AAAAAAAG TCGAGGGGG GCGCGTACC 1140

60

5

(2) INFORMATION FOR SEQ ID NO: 243:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

AUCCACAGC TCTGTGCGA GGGGCTGAG GCGAGAGGC CCGGAGGCG CAGTTGCGAA 60
CATGTGCTAG AGCAGAGAG GCGGAAACCC GTTGGCCGAG CCGAGCGAGC TTGACAAACC 120
CTTTCACCA CCACAGCCT ATGAGCCTCC AGCGCTGCC CGATTGCTC CAGCCCTCAC 180
TCCCTCTTG CAGCCTCTGA GAAAGCTCAG CCGCAGAAA CTTAAGAACT ATGCTCTATA 240
CAGCACTCAG GCGTCACTG CAGCAGCCAC AGCTGAGCTG CTGAGAGAAC AGGAGAGCT 300
GAAAGCGAAG GAGAGAGAT TGGACCTAAG GAGAGAGAG CTGAGCAGTG CTGCGCTGG 360
RGGAGAGCT ACTGACAGA ACAAATGGCC CCGTCTACT TCTTTTGTCT CAGTTGAGCC 420
CTGCTTTTTC CAGGAGCTT CGATGAGAT CCGCCAGAA TTTCAGAGA CTGTATCCAC 480
CATGTACTAC CTCTGATGT GCAGCAGGCT GGTCTCTCTC CTGAACTTCC TCGCTGCTCT 540
GGCGAGTTC TGTGTGAAA CGAACATGG CGCAGGCTTT GGGCTTTCTA TCTCTGCTCT 600
CTCTCTTTTC ACTGCTGCT CTTTGTCTG CTGTATCCGC TTGGGATTC AGTGGCTGGA TCTCTGCTCT 660
GAGTGAAGT TCTTCTAAT TCTTCTTTT CTCTCTCAT TCTTTCTGTC AGGATGTGCT 720
CTTTTCTCTC CAGGCGAATG GTATCCGAG TTGGGATTC AGTGGCTGGA TCTCTGCTCT 780
GTTGTGTCCG AAGCGACAC AGCATATTC GTCTCATTC TGTGTGTGTC CCGTCTCTTC 840
ACTGGCATTG CTGTGCTAGG AATGTCTATG CTGAAAGGGA TCGACTCTTT ATACGCGCC 900
ACAGTGTCCA GCTTTGAGA GCGCCAGGAA GATTTGCTG CTGTGTCTTT CTCCAGCCT 960
GGGTTGCGA CCGCACTTG CGATGCGAG CCGTGGGCTT GCTGAATATG CCGTTGCGGC 1020
CCGTTAGCC CTGACTGGA TCGCCTGACC CTGCTACTTG AGGAGCTGA CTTAGCTCCC 1080
GTCCCTTAGG TCTCTGAGC TTGAGAGAGC ATCACTAAT GATGCTCTCT CCGTATGCT 1140
CGCAATCTTA TGGCATGAC TGTGAGACT GACAGGCTG TGGGAGTTT ACTGTGACT 1200
AGTCCGCCCA TCGAGCCACA CTCTGCGAC CTCTCAGAG CCGCAGCCA GCTTCTCTCT 1260
GCTGTGAGC GGTCTGCTCT TGGTTATTT AATTAAGAG AAGTGGAGC TCGAAGAAA 1320
AAAAAAAAA AAAAAAAG GGGGTCGC 1350

5 (2) INFORMATION FOR SEQ ID NO: 244:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1529 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(2) INFORMATION FOR SEQ ID NO: 245:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1537 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

15 TCCGAGAGCC CGGGGGTTC CAGCTCTGCC TGTAGCAGAG CCGTGTAGAG GAGGAGAGAG 60
AGGATGTGCT GAAATAGCTC CGGAGATCT TTTTCAGCTA GGGCATMAAC TGTGCACTGA 120
ACTGTCTGCC GAGAGCAGCT GAGGAGAGCC TGAGCTTCCA CTGGTGTCTC TGGGGCCGAC 180
GCCCTGTGGG AATGGGGCTCT CTGTGCTCTT ACCTTTGTGC CTTCCTTGGC CTGGCAGATT 240
CAGCTCAGCC CAGAGGCCCC TGGAGCTCC GGGCCCTTGG GTGCCCTTCT GAGTGTGGGG 300
AAGGCGAGAC TCAGATGAG ATCCCATTTG ACTCCCTCTG TATGTACTGT GCGCTCTGCT 360
GGCTCTTGAG GCTCTGAGAT CCGATTTGTC TGTGTTAGTC AGTGACACAG TTCCAGGCGA 420
AATATATCTA TGTGTTGGTC CAATCTACTG GAGCCAGAGA GACAGTACTT TGCAGAGAAA 480
AGGATCACTG CAGAGCTGAC TGGATTTGCT AGATTTTATG CCGGATGATC TCTGCTGAG 540
GAGGAGGCTT GTTTTCAAAA TAGTTTCCAT CATGAGCTTA TCATGAGCTT CCGACCTCTC 600
CAGCCAGGCT AGAAGCGAAA CAGAGTGGCC AGATTTCTCT GCGCTGTGTC GAGGTTTGAG 660
GCCAGATGCT ATTAGACTGT AAGCCAGACA GCGCTCTCTC GCGAGCTGTC TACCTTGGCT 720
TCAGCTGTAC CTTCGTCTCC GGGGCTGAG CTATTAAGCA AAGGGGACAG GAGGCCAGAA 780
GAGGACACTG GGCAGAGAGA TCAGACAGA GTACATGTCT CTGCTCTCTT TTTGAGTGTG 840
GCTTTGAGCA GGAATATATG AATTAATCAC TCCCTATCAG GTTTTCCGAT ACAAAGTGC 900
TAGAATAATAC ACAATATTC CCAATGCTTA AGTTGTGCTA ATGTCTTTTC AGTTCTGGG 960
TTGGGAGTGG GAGGTTGCGA GCGTTGTGTT GTGCGAACCC GTGAGTGTCT GTTTCACAGG 1020
AGGATTTTGA GTTCTCCAGG GTCTCATCAT GCGAGTGATT TGTGAGCGGA GCGCTCTGCC 1080
CTGCTCTGCT TCAAGTCCAG GAGAGCTTTG AAGCAGTCTA GCGTTGTGTT TGTACCCGAT 1140
GTGTCTGTGC TTTGTGTGAT CACTCAGAGA TCAGCTCTGG ACCTCTGGGG TTGAGTGTCC 1200
AGTATATGCT TATGCGGGCC CACTCACTAT GGTGCGGCTGA GTGGAAGTTC CTTAACCATG 1260
TCCCGCAGAG CACTGAGGTG CTGCGCTTTT TAAATGTCTC GTTTGTGCCC GTAACTTCTT 1320
TGTAGTGTCT CATTTTGGCA TTGGCAATT CAGCCTGGAA GTCTGGGGCC ATGACAGCAA 1380

15 TCACTCTCTC CCGACCTCTC TGAGCTTAGA GGAATATTTG CTCAGATCCA TTATTTAAG 1440
CAGGAATTTG TGTGCAATG AGCTGCATGG TTTAGCGAGT CTTTGGGAGC CTTGGAGTGC 1500
5 CTCAGAGACA ACAAATCTTG TACTAGAA 1529

20 GTGCGAGTTC CCGCCAGGCC CCGAGCGGCC TTCCGCGGCC GGGGGCGCTCC CAGAGCAAAC 60
GAGGCGCTTG AGAGCTCCAC CTAGTTTACA GZATTAATTC CCGAGCGACA ACTCGAGTTC 120
ACCAATGCTT AAGGCCCGAG TGTGTTAGC TGTGTTATTA ATGTCTAGC TGTCTGTGAA 180
TGGCCCTGAA TTTTACCTTT CAGGTTATTC TTCCAGTTAC ACAAAATCT ATGAGGATGG 240
TTGTGAGAT TATCTTACTC TATCGATTA TGTTCAGGAT TTTTGTATTC ATCTTACAAA 300
GCAAGCTGCG AGTTTGTAAA CTGAATTTGA ACAGTTTGGC GAGACCTCTA ATGCTGTGTT 360
TACAGCAGAT GATGCTTTGC AGAGCTTGT GAGACTCATC TATCGACAGG CACATCTAT 420
CCCAATTTTC TCTTATATGG GAGCTGCTCT GTGTAAATAC CTGTCCCATC ATCTGACAT 480
TAGCCGACAG AGTGCGACT TCGCGAATTT CTTACTTTCA AGATGTGGA CTGAATATGA 540
AGTTAAAGAT CAGCTGCGAA AAGGGGATGA AGTTACTGGA AAGCGATTTT ATGCAATTTT 600
ACTGTTTCTG GCGAGACTTT ATCTTACGCT GAGATCAGG GGAACAAATG GAGAGGTTAC 660
AAGAGCAGAT ATTCTTCAAG TTGCTCTTGG AGAATTTCTG AATGCGCTGT TTTCTAATCC 720
TATGATGAC AATTTAATTT GTGCAATTA AATGTTAAG TTGACAGAT CAGTTTGTGA 780
AGATGCTTGG AAGGAAAGAG GAAGATGA TATGGAGAA ATTATTGGA GAATTCAGAA 840
CTTTGTGCTA GATCGAATCT GCGATGAGA TGTAAACAG ATGCTCTCTA AGCTTGTAGA 900
50 ACTGCGTCA AGTAAGCTGG GCGAGTCTCA TCGAACTTCA ACATATAGAG AAGCAAGCC 960
AGAAATGAT CTTAATCTACT TTATGATGA ACCAACATTT TATGATCTG ATGCTGTCC 1020
TTTCACTGCA GCTCATCCAG ATTACAGGA GAATTAACCA GAATTTACTTG AAGAGAGGA 1080
CTTTTGTCCA GATTTGTGAG AATTTGACAG AGATTTATCC GGGGCTGGTG ATGCAATCTT 1140
GATGATATTT GATGATGAGA TGGAGCCAGA GATGAGAGAA GCTTATGAAA AGTTTGTGTT 1200
60 GGAATCAGAG CTTAAGCGAA AAGCTTAAG TTAATTTTCA GCAATATGAT TTTAAGAGC 1260

AGTTTAGSTA TGGTCAATTA GCGAGACA AGAGCCAG AAATGTGTC ACATCTATAC 1370
CAATTTAGG ATGTGAGT ATGTACTA TGTATGCAAC TTATATTTTG TTATACACTA 1380
TCTGCCAAA TAACTTTAT TCTCTATAC TTAAATGTC TAAATATAT TAAATATTTA 1440
TTATATACAG TTATATCTAC TGTATTGCT GCAATAAAT CAATTTTGA ATAAATPAAA 1500
AAAAAAAA AAGGAGGCC GCTCTAGAG ANCCAG 1537

5
10
15
20
25
30
35
40

(1) INFORMATION FOR SEQ ID NO: 246:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 506 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

TCGAGATT GCGAGACC CCGCGGGT GCGTTGCTA TCGTTTCCA GAACTACTC 60
AGCGGCCAG CTCAGAGAG TTGAGGAAA GTGCTGTCC TCGCTCTCA GAGCGATCG 120
ATAGGTCCA GCGGAAATA AATCATGCC CTTTCTGCT CAGTGTGAAA GCGCACTCA 180
AGATCTGGC GCTGATATT ATCACTCAC TGTATACAC AGTATCTAG CTATCTGAT 240
CTGTGTGCG ACTGATACA GAATACACA CATGTACAT TGTGTGAGG GTGTTTGAC 300
TTGTGAGCC AGTATGCTCT CTTCGGAGC GCGCGCTTAT TTACCGGAG CTTCTGTCA 360
ATCCAGCGG TCTTACAG AAAAGCCTG TGCATGAAA AAAGAAATT TTGTATTTT 420
ATATTACTTT TTAGTTTAT ACTATGATT AACTATTTT CTGATTTAT CCAAAAAAA 480
AAAAAAAA AAAAAAATT TGTGG 506

TTCTCTCAG GCGAGGTGC CTTGAGCAC TGGTAGTTCT GGGCTGGGA GGGAGAGGG 240
CTCGGCTTT CTCTGAATG AACTGTCTC TTGACAGTT CAAGTACTTG TTCTCAAAAC 300
ATTTCTAAT TGAATGGAG GTTTCTATA GCATGTGTTT TTTAAGGCAT GGAAGGAAA 360
GAATGCTAA GCAATGATG TTTGTTTCA GTGGATGGG CCGGCTTCT CACTGTGCG 420
GGCTTCCCT TGCATGTGCG ACCTTTGCG AGGGCAGCA GCGAGACTT TCCAGCTTC 480
TCCACTGAA GCAACAGGG GCTTGAGCG TAAITTGCT AATCAGGCG ATTTTTTTG 540
TCTAGTATC TTTCACACT GTCAACCTT CTATTTTTT TAAAGTTCT GTTCTGTGA 600
TTAACAGAA ACTAGAGAGA AATGTTTCT GAGGCAAGT TATTTGTAG ATCCCAAGG 660
GAGGTTGCG TAGAGAAAA TAGTAGCTG GTTTAGAAAC TGACAGGCG AAGAGCCAG 720
GAGCATGCG AGAGGAATTT GCGAAGATC TACCTGAGA TACGCTCTT CCAATGCTT 780
CACCACTGA ATACACGCG CTCMAAGTG TTTTCTGCT TTGAAAAAA AATTTCCA 840
AGCTTTTAA GGTGCATTTA AGATCATG TGACTTTAGA ATGGAATGCG CGGCTGTGC 900
AATCTGTACG TGTCTAGAA GTTTCAATC CTCTGGAATG CATGTGATAC TCAATCTCAT 960
TTTGTGCTT TGAITGCAIT TTTGTCTTT TACGAGTCT GTCCCTGTG GTGCTGTCTA 1020
AGATGTGGA CACTGTGTT TTTGTATAG ATTGAGCTG GAGCTGCA TCAAGTCTT 1080
TATATGAAA TTAGGACAA CCAATCTGT GTCCCTGCT TGTGTGCTAA TGAATGAGG 1140
GAGGAGGG ATGTCAAGCC AAAGTAGCG CTTCCCATG GCTTTGCGCA GCGCAGAC 1200
TTCAATGCT TTACATGTT CTGTGTAAT TTAAAGTTTA TGTGTATAA CCGAGCTGT 1260
TTCTGTGAAA CTGTATATT TGTATATAA TATATGCTA CTTTGAGAA AAAAAAAA 1320
AAAAACTGA GGGGGGCGG GTACCCA 1348

5
10
15
20
25
30
35
40

45 (2) INFORMATION FOR SEQ ID NO: 248:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1766 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

55 GTCCGAAAT GCGAGAGCG CAGCAGCGC CAGCAGACA GCGGAGTAA AGCAGTTCA 60
GCGAGCAAT TCGCGAATA TTCTATTTCC CTTCTCTCTC TCCCGCCCG TATCTCTTTT 120
CACCTTTCT CAGCTTCC TCGGTATCA TCGCGAGCG TCGCGGACA CTCAGTCCA 180

60

5 TTCTCATCTCC TCGTCTGCTCT TCGAGAGCCGA GCGCTCTGCGG CCGCGCGCGG GCGCGAGGCC
AGAGGCTGCG CCGCGCGCTCG GAGAGGAGAG CTGACAGCTCC TCGTCTGCGG TCGACAGATCT
GATTCTCTCG AGAGATGTGA AGAGAGCTCG GTTTGTCTTT GCGACAGCGC TCGATCATCT
GCTTTCTCG GAGCTTTCTA GTGTCTATGAG TGTGTTTTCT TACCTCATGCG TGGCTCTCT
CTCTCTGAGC ATCAGCTTCA GGAATCTACA GTGCTGTATC CAGCTGTAC AGAGTCTAGA
10 AGAGGCGCAT CAGTCAAG CTTACTGTGA GTTAGAGATCT ACTCTGTCTCT CAGAGCTTT
CGATTAATC ATGATCTGCG CGATGTGCA CATCAGACAG GCGCTGAGC TGTATTGCG
15 TCTCTCTGCG GTAGAGATCT TGTGTATCT CTGTAGAGCTG GCTCTCTCTA TGTGCTGAT
GAGCTATGTT GGTGCTGTTT TTAGCGAT CAGCTTCTTA ATTCTTCTG ACTCTCTAT
70 TTCTAGTCTC CGATGTCTCT ATGAGAGTGA CAGAGCGCAG ATTGATCATCT ATGTTGCGAT
CGCGCGAGT CAGACCATCT CATTGTGGA AAGATCTCA GCGAACTCTC CTGAGTCTC
80 CAAAAAAG GAGATATGAG TACATGGA CAGAGATCTC AAGATCTCT AAGAGCTAT
25 TTAATATGTA TACGTCTGTT ACTGTGACTA TGAAGGAAA TACTCATGTT CAGCTGAGC
CTCATATCCA AGCTTTTTTT TTAATTTGTT GTTTCTCTCC ATCTTTTCCC TTTTACCTCT
100 AGTATCAAGC ACGAAATGTT ATGAGCTGAT AAGAGACTTA TCTTAGACTCT CAGAGAGGA
1080 AAGATCAAA TTTATAGAT AAGTCAATAC GTTAATGTTG GTAGAGCTTT TACCTGTAGC
1140 TGAAGAGCG AAGATATGGA GTTAGAGAG AAGATGAGAG AAGAGCTCTG GTTCTCTCTG
1200 TCGAGTTTTC AGGACTATCT TTACTCATCT ATCATATTA GTTTTGTGCT TAAAGATGCA
35 TGAATTAAT ATGAAATAT TATTTGCGA CAGGAGCTCG ATAGCTTCTT TGGTTTTTTT
1320 TCGAGGCTCT CAAATCTGAT CTCTCTGCCC CAGATGTGA CAGAGTACCC CTGATATCTC
1380 TTTTCTTTAA TCAATTAATCT ATCAACTTGA TAAATTAATCT ATAGGTGATTA GTGTATATTC
1440 CTGATTTCAA GAATGCGATC TGTATAAAAA GATTAAGAAAT GGAAGATGCG ACTGAGAGCG
45 AGTCAGAGCG CATGCTGCG TGGCGGTGAC TCGCTCTGCC ACTATGCCCA GCGAGAGGAA
1560 RCGTCCGCCA TTTGCGAAG TGTGTTTAC GTCACTGAGC ACCGTTCTG AGCTTAATTT
1620 TGAGAGCTCG TTCCGGAATG TGTGTTCTCT CTTCTGCTCT CCGCAGCTCA AGTTATATTA
1680 ATAAAGTTGT ACTTTTCTTA CTATATAATA AAAAAAATA AACTGAGAGG GCGCGCGGTA
1740 CCGAAATGCG CGATATGAT CTGAAA
1766

(2) INFORMATION FOR SEQ ID NO: 249:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2664 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

5 AGTGTCTCTG GAGCAGCGCG AGTAAGGGA CTGAGCGAG CAGATTGCGG GATTATTTCTA 60
10 TTTTCCCTCTC CTCTCTGCGG CCGCTATCTCT CTTTTCAGCC TTCTCCAGCC CTGCTGCGG 120
TASCTGCGG GAGCTGTGCG GCGCTCTGAG TCGCATTTCA AATCTCTCTC CTGCTTCTGA 180
GCGAGCGCT CCGCGCGCGG CCGCGCGCGG AGCGCGAG GGTGCGCGG CCGCGCGG 240
AAGAGCTGCA GCTCTCTCTG TCGGTGTGAC GATCTGATTT TCTGTAGAGA TGTAGAGAG 300
ACTGCTTTTG TCTTTGCGAC CAGCTGATC ATGCTCTCTT CCGTGTGCGC TTTCTAGTCTC 360
20 ATCATGTTG TTTCTTACTCT CATCTGCTCT CTCTCTCTCT TCGACATGAG CTTGAGATC 420
TACAGTCTCG TCACTCAAGC TGTACAGAG TCGAGAGAG GCGATCATCT CAAAGCTCTAC 480
CTGAGCTGAG AATTAATCTCT GTTCTGAGA GCTTTCTGTA ATTACATGAA TCGTCTCATG 540
25 GTCACTATCA ACGGCGCTCT GAAATCATTT ATGCTCTCTT TTTCTGTAGA AGATCTGCTT 600
GACTCTCTTA AGCTGCTCTCT CTTCATGTCG CTGATGACTT ATGTTGTGTCG TGTTTTAAAC 660
30 GGAATCAAGC TTTCTATCTCT TCGTGAATCT GTCATTTTCA GTCTCCGAT TGTCTATGAG 720
AAGTATGAGA CCGATTTGTA TCACTATGTT GCGATGCGCC GAGTCTGAGC CAAATCAATT 780
35 GTTGAAAGCA TCGAGCGAA ACTCCCTGGA ATGCGAATA AAGAGCGAGA ATAGTATCAT 840
GGAAGCGUA AATGAGCAG TTAATAAAC ACGATTAAT AGTTATTAAG TCGTTACTCTG 900
TACTATTAAG GAAATTAATCT AGTGTCACT TGAAGCTGCA TTTCAAGCTT TTTTATTAT 960
40 TTTGTTTTTT CTCCGATCTT TCGCTTTTAA CCGTCAATAT CAAAGCAAAA AATTGATGGA 1020
CTGATTAAG AACTATCTTA GAACTGAGA GAGGAGAA TCAATTTAT AGATTAATCT 1080
45 AATACCTTAA TGTGTGTAGA GCTTTACTCT GTAGCTTGA AAGGAGAAAGA TTTGAGGTAA 1140
GAGAGAAAT GAAAGAGAC CTCTGCTGCC TTTCTGTGCG TTTTCAAGC TAGTCTTACT 1200
CAGCTATCCA TTAATGTTTT GCGCTTGAAG AGTCAATAT AACTATGAA AAAAAATTT 1260
50 GCGAGAGGA GTGTGATACC TTTCTGTGTT TTTTGTGCA GCGCTCAAT CCAATCTCTC 1320
TCCCGCAGA TGTGAGGAG TACCGCTGAT ACTCTCTTTT TTTAATGAT TAACTATCAA 1380
CTTGATTAAT AACTATGAG TGTATGAT AATCTCTGAT TCGAGAGT CCAATCTGTA 1440
55 AAAAAAGATA GAAATGAAA GTGAGACTGA GAGGAGTCA GCGAGATCT TCGCTGCGG 1500
GTCACTCTCT CTGCACTAT CCGCAGGGA GGAAGCTC GCGATTTGCG GAAAGTCTTT 1560
60 TCTAGCTCAC TGAACAGCGG TTTCTGAGAT TATGTTTGA AACTCTTTCC GATTTCTCTT 1620

5 TCTGCTCTT CCGTGGCCA CTTCAAGTTT AATTAAATAG GTTGTACTTT TCTTACTATA 1680
AATTAAATGT CTGTACTGCT TGTGCACTGCT TGTAACTGTT TTAGAGAAAA AATTAACTGT 1740
CATGTGGGCT CTTCAAGTTT TGAATTTTGT TGAUCCATCT TCAUATCTGG GGGGAACTAT 1800
CTCAAGAGGT GAATTAACAGA AAGCTTTTGT TCTTTGATCT TTTCGCCAGA TTCAAAATCTC 1860
CGATTCOCAT TTGGGGCCA GTTTTTTGT TCAUCTTCAA TATGAGATTT CAGCGAATCT 1920
GAAGAGAAAA TCAATCTGTA GTTCTTTGAG GTTCTCACTC ATGATCTATGA TCTTTCCAGG 1980
GGAAATATCA CTGGCGAGTT TAACTTAAGG GCTATGATAT TTGATGCTCC CAAGTTHAAG 2040
CAGCTCCAAA AAGTAAATGA AGGAATTTGT CTAGCTGTCT TGGAAAAATT AGTTAAGGAT 2100
TTGATGGGCT AAAGATAC CTTGCTTTAC TCAATCTTAT TTCTTTAGCC CCGTTTGAT 2160
GTTTTAACTC GTTTCATGTC CTAGTAGAA GTGCATCTC CATCTCATC CTCTGCGCTC 2220
CGAGAACTC AGTATGTC TTTTGGGCT TCGCTCCAA AGGAATCTCT GCAATGGAG 2280
TGCATATCC AGTCTTTTC TTTTGTGCT GCTGTGTTTA GATTAATGAA GAGATCTTGT 2340
TCCACACAG GATTTTATTT TTTTAAAGA AAACCTATA GATGAAATAT TACTAATGAA 2400
ACTGTGTGTA CGTGTCTGTC CTTGCAATAT AAATTAACAG TAGCACTTAA GAGCTTGA 2460
TCTTGTCTTC TGTAAATTTT CAATTCATG TGTATTTAT TAAATAAAAA AAACAAATA 2520
AAAAAATAA AAAGAGGCGG CCGCTCTAGA GATTCAGCT TTAGCTAACC GTGCATGCA 2580
CCTCATAGC TCTTTCTATA GGGTCTCCG AAATTCATTT CAGCGGCGG TCGGTTTTAT 2640
AAGGTGTGTC ATGCGGGAAA AACC 2664

(2) INFORMATION FOR SEQ ID NO: 250:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 865 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

50 CTTGGAGTC AGTAAACAGA TTGAGCCAT TTGGCCCGCA GCGCTCTTGT CTGGAATCC 60
GGTGTCTGCG GATTAAGATC CCGTTCTTA AGGTGGGAT CCGTGTCTC GGGATGAGAT 120
TTGGGCTTTC CTGCGGCTT TGGTGGATC GGTATCTCA GATGAGATTT TTAGGTTTCC 180
TGGGGGCTTT CCGGATCTC ACTTAATAT CCGACTCCAA GATGAGGAA GCGCGGAACC 240
TAGAGGCTT GATTAAATG GTCACTTAC TGGTCTGTC AGTGTCTCG GCGATGCAA 300

5 TGTGGTAG CTTGCTCTA GGTGTCTGCT TTTTGGAG GCTTCCCGCA CATACTTGT 360
GACTAGTCA GAGAACTC TTCCCTTCT ACTTACAT CTCAATGGC TGTGCTTCA 420
TCAACTCTG CATGTGCTT TCAAGCATG GTTGGGCTA GCTCAATATC TGGAGAGCA 480
GCCAGCTTAA CTTGCTGTC CTGAGCTTAA CCGTGGCCAC TGTAAAGGCC GGTGTGCTG 540
AAGCCCGAC CAGAGTGGC ATGTGGGCC TGTAAACCTT GAGAAAGGAG CAGAGCTGTG 600
GTGGAGGT ACCAGGACC CACCAAGATC CCGATCTCTA CCGCAAGCTC CAGAGAGAG 660
ACCCCAAGTA CAGTGTCTC CCGCAATAT TCTTCCGCTA CCAATGGGCT TCTCTCTT 720
GCATCTGG CTTGCTCTG AGCAATGGC TGTGCTGCT TGGCTTGGC CTGGAATTA 780
GAGGCTCTA GCAATGGCC TCAATCTTA TAAATCTCT TCAAGAAAA AAAAAAAA 840
AAACTGAG GGGGCTGGT ACCCA 865

(2) INFORMATION FOR SEQ ID NO: 251:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2082 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

35 TGGGGGGG AATGGTCTC TGGCTGAG ATTCGCAAT CTGGAATTC TCAATACTT 60
GCTACTGCT TTTTATTTT TTTTATACA CCGCCCGGCC CACCCCGCG ACTTGCAGA 120
TGTTCATGA TCTGAGAGA GTTCTCATG TGAAGCTTG ATCACTTTT AAGCTTCAT 180
CATTCATTA TTTTATCTC TTCTGCGCT TCAATCATG AACTGTGTT CATTTTCTT 240
CTGTGTGCT GTTTTATTT GTTGGATTT TTTTATTTA TTTTACTTT AAGCTTCT 300
GTGTGCGCA CTTTATTTT AAGCTCCAC CTCACTGCT CTCAACCAT CTCTTCCAG 360
ATCAAGAAA AAAAAAGCA AAGTTTATTT TCTTCTCTT GAGTTCTCA TGTGAGATT 420
AGCTTCAA GCAAAAAA ATGTGAATG TTATAGACTT GCAAGCTGCC GAGTTCTATC 480
GGTTTTTT TTAGCATG TTATCTTAA ATAGAGAAA AATGCTCAT GAACCTTCA 540
CAATCAACC TCAATCAAC TTCTGCTGT GACTGTGAG TTTTGGCTT GTATGCTCA 600
ATCTAGAGT TTAGTCTGCC ATTAATAAA CTCATTTCTA TCTCATCAT TATTATGCTT 660
GCTACTTGT CTTAAGACA ATGAACTTA ACTGTTTCTA AGACTTTATG GAAGAGAGC 720
ATTATTTAA TAAAAAAA AAGCTTCAT GCTGGAGTG TATGATTA TTTATTTTC 780
CTTTTTTT CTTTGGCT TGGAAATGA CTTTCAAGA CTTATAGAT GCAATCTA 840

5 CTTTGTGTTT ATTGCTCAAT GACTTTTGTG AGTTTAGAAC AAAACAGTGC ACCCTTAGAG 900
CCTTCTGCC ATGAAATTTT GCKTCTGCTC CAAACTGCT TTGAGTTACT CAGAACTTCA 960
ACCTCCCAAT GCKTGBAGG CATTCTCTGT GCAAGATAC CAGATGGGT TAGACATTTA 1020
ACCTGGCAA CATTGAGAA CTTTATGAT TTCTTTTATA ATAGATACA GCGCCCACTT 1080
10 TGGGCACTTA ATTGTGCTA TTGCGGAGAA GAGTCTTAAA ATTATTTT TTAAAAAGAA 1140
AACTGCCCC TTATTTTGG TTGTTTTAT TTATTTTATA TATTTTGTG CTTTGTGCA 1200
15 TTGTCAATG TGGATGCTC TGGTTTCTTA GTATATATAT TAAATCTAGT TTATTAATC 1260
TGTTAGCCCA GTTAAATGAT ATGCTACAGA TAAAGGATG TTATAGATTA ATTGAAAGA 1320
GTTAGCTCTG TTATGCTGTA GAATTTTATA AGATGATG CACTAAATG TTACTATTTG 1380
20 TGAATTAAG GCGGTAGAG TTGCAAGG GACTGTTTAA AAAAGTAC TTATACAGA 1440
TGTGTTGCA ACTTAATAT AGTTGGGTA TGTGTAGTCT TTGCTATACC ACTGACTGTA 1500
25 TTGAAMACCA AGATATTAAG AGCGGAAGG CGCTGTGTTA TATCTGTAGG GGTATTTTAC 1560
ATTCAAAAT GTATGTTTTT TTTTCTTTC AAATTAAG TATTTGGGAC TGATTTGCAC 1620
TAAGATATA CTTGCAACA TATATACAA AAAAATG CAAACTGTT TAGAAGCTTA 1680
30 ATAAATTTA TCGATTTATA AAATGGGAT TACTGCAAG TTATAGATG ATGCAGATT 1740
TTTACAGTT GTATGTGAT GCAAGCTG ATTTCTGTA ACTTAAMAA AAATCCGAG 1800
TTTAAAGCC ATATATGAT AAATGATAT TTCAAGGACT GACATCTGT CTTTAAAG 1860
35 AAATGAAG TAAATCTTAC CACATTAAT ATAAATAT CTTGTGATT ACTTTCTTT 1920
TACATATTT GCTGTGAAA ATGTGTTAT ATCTTGATG ACTACTTAC CACCGTGTT 1980
40 GTTCTATGT GCTTTTCTT CATTTCAT TCTGTTATA TCAAGAAAG AATAATCTAC 2040
AATAATAAC GCAATTTTT TTGAAAAA AAAAAAAAA AA 2082

(2) INFORMATION FOR SEQ ID NO: 252:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1482 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

50 CAGGAGGCT GCGCCGCGG ACTTCTCT GCGCTCTC CTCTGAGG CTCCGCGTT 60
GCGCCCTCG CCGCTAGGA GTCTTAGC AGGATGAGG CTGTTGTGA CTTTACCAA 120

5 GAGGTGATGA AGCAGCGAGA TCCCGGATC CAGGGCTACC CTCTGATGG GTCCCGCTTG 180
CTAATGACT CAACTCTCT GACTAGCTG TACTTCTGTC TCTGCTGCG GCTTGGATC 240
ATGGCTAATC GGAAGCCCTT CAGCTCTCT GCTTCTGATA TTCTTACAA CTCTCACTG 300
GTGCACTCT CCTCTGATAT TTCTATAGAG TTCTGATGAT GCGCTGCTT GAGGACCTAT 360
10 ACTTGGCCTT GTGACCTCTT GCACTATTTCC AACAGCCCTG AGGCACTTAG GATGATTTGG 420
GTGCGCTGCG TCTTCTCTTT CTTCAAGTTC ATTGAGCTGA TGGACAGAT GATCTTATTT 480
CTCCGAAGA AAGAGCGGCA GTTGACCTTC CTACATGCTT TCCATCACTC TGTGCTTCCC 540
15 TCGAGCTGCT GTTGGGCGGT AAGATTTGCC CCGGAGGAA TGGCTCTTT CCAATCCATG 600
ATAAACTCTT CCGTGCATGT CATTAATATC CTGTACTTAG GATTAATCTC CTTTGGCCTT 660
GTGCGACAC CTTACTTTTG GTGGAAGAG CACATGACAG CAACTGAGT GATTCAGTTT 720
20 GTCTGCTGCT CACTGCAAT CTCCAGTAC TACTTTATGT CCAAGCTTAA CTACCAATAC 780
CAGTCAATTA TTGACTCAT CTGGATGAT GCAACATCT TCTTCAAGT GTTCTCCAC 840
25 TTCTGTATC ACTCTTATC CAAGGCAAG CGCTGCGCC GTGCACTTA CCAAAATGGA 900
GCTTCAGGTA TTGCAAGGT CAGGCGCAC TGAGAGGAT GCTCTAGATA GCGCGCCAC 960
TAAGTGCCT AGGACTGAC CTTAGGCGAG TGTCTGTCAG TCGCTCTCTC ACTTACACT 1020
30 GTGACCAAG CTTATGTGAT CAGGACTGAG CAGGCGACTG GCTCTGCTCT CCGCAGAGCT 1080
GCTTACAGG GACACAGGCT TTGTTTCTC ACCCACTTC CCGGCGCAC TCGAGGATG 1140
35 TCGCTCATT GCTGTCTCC ACTCCAGAG TGGGCGCTTA AGGCTCTTA CAGTTATTTT 1200
CGCTTCTCT CTTTAAACT TGGGAGGAA GCACTCAGG CTTGCGCCAC AAAGGCTCTC 1260
GTGCGCTTTT TCTCAGACA GAGAGGTCA GCAATATAT CACTGTGAC CAACTCTCAC 1320
40 TCTCCACCC CACACACTGA AGCAGTAGT TCTGCGGCAA AGGTCAGGAT GCGCGCGGCG 1380
CTGGGAATAC AGCTGTGGA GCTGCTTAC TCACTCTGT TCTTATTTA AGTGAACGA 1440
45 GGAAMCCAA AAAAAAAAA AAAAACTGA GCGCGCGCGG TA 1482

(2) INFORMATION FOR SEQ ID NO: 253:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 834 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

50 GCGAGAGCG CGTTGCGCG CTTGCGGCTT ACGAGTCTT TAGCCAGAT GAGAGGCTTT 60

5 GTGAACCTCTT ACAGAGGCT GATGAGGAC GCGAATCCCT GATTCAGGG CTACCTCTCTG 120
ATGGGGTCCC CCTTGTCTAT GAGCTCCANT CTCTGTACTT AGCTGTACTT CTCTCTCTCA 180
CTTGGGGCTC GATCAGTGGC TATGTCGAG CCCCTTCAGC TCGTGGCTT CATGATCTCT 240
TACAACTCTT CACTGTGGC ACTCTGGCTC TACATCTCTT ATGAGTTCTT GATCTGGGGC 300
TGGCTGAGCA CTTATACCTG GCGCTGTGAC CTTGAGACT GCACTTAGG GCACTGTCCG 360
TCAGTGGCTT CTTCCATAC ACTGTGACC AAGGCTTAGT TGGTCAGAC TCGAGAGGG 420
ACTGGGGCTT CCGTCCGAC AGCTGTCTTA CAGGGACAC GCGTTTGGTT CTTGACCCAC 480
TTGCGCGGGG CAGCTCCAGG GATGTGGCTT CATTCTCTCT TCCACTTCA GAGCTGGGG 540
CTAAMAGGGC TGTACAGTTA TTTCGGCTCT CTTGGCTTAA AACTTGGAG AGGAGCATC 600
AGGGCTGGCC CCGACAGGG TCTGTGGCC TTTTCTCTCA CACGAGAG GTGAGCATTA 660
ATGTCTACTT GAGCCGATC TCACTCTCTC ACCGACACA CTGAGCAT AGCTTCTGG 720
CCAAAGTCA GGTGGGGG GGGCTGGGA ATACAGCTTG TGGAGCTCT TTACTGACT 780
TGTCTCTTAA TTAAAGTGA CAGAGAAC CAGGAAAA AAAAAAAAA AAAA 834

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 254:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

TTGAACCTTT AATTTTGG ATCGCAAC TCTAGATCC TAGATGGAA GCTGTCTCTC 60
ATTCTTCAT GCTACCTTC CAGATCCAG GAGATGGTCA AGAGCTGCA CCGCGAAC 120
CCAGCAAGCT TGGATGGA CCTCTCAT GAGCTTCTG AGACTTTTG CAGATCTCC 180
AAGTGTCTG GTCCGAGC CTTTCTCTT GTCTGCGA CTGTCTGA GGAATGCTCC 240
CTTCCCATG ACGATTTC TGTCTCTCT ATGCGGATC TCTTACTCT TACCAATGAC 300
AGGTTCTCTA AGTGGCAT GCTCTTCA AGACATTA GCAAACTCT ACTAGAAAA 360
GACTATTTCT TGGCTCTG CAGTCCAC CAGAGGCTG TGGAGCAGC CATATGGCT 420
CTTCAGATG ACCGTGAG GATGTGAG TATTTTCAA GCATCCAGC TGGCATTACC 480
AATATCTCG AAGTCCCT GAGCAGAG TCTCATACT ACTAGAGGC TTGATCTCG 540
GTGTCTTCC TGTCTCAT AGAGCGAGG TTCATGGG ATTGCGAG CATGTGACT 600

GGATAGCTT TCGGGAGG AGAGACTTC CTCTCTGG GACTTCATG CAGGTGCGAG 660
TTGCTTACAC CAAATACAG GAAATTCAG ATCGAAGGA AGTACAGTA AACACTATTA 720
TCTATCTTG ACTTAAAGG GAAATAAATTT CTGAGAGAT TATATTTGTC ACCGAGCCT 780
TAAATCTTTC TGTCTCTCTG ACTGATGAA ACTTGATTTG GAGAGCATT TTCTTATGG 840
AAGGATGAG ATTCCGAGAG ACCTGCAATG CTTTCTCTCT GTTTTATTTA ACATCGACA 900
AATGAAATTC TTACAGCTTG AAGGCGAGG TGTGCGCAGA TGTGAAGAG ACCTTCACTA 960
TCAGCCCTAA CTCTCTCTC CCGAGAGA CTGTGTGGG TCTGTGGCA GCTGTGAGC 1020
CGAGCTCTGT GTTGAAATG TTGTGTGACT GTCAAAATG GAAAGAGGG GTTTTACAT 1080
CTCTTAAGG ACCTGATCC AACACAGTA GAAATGACTT AAACCTTTAA GCGCAGATA 1140
TTCTGTACA GATTACAGA ATGTGTCTG ATGTCTCTG TCTGATTTT TCACTCTGCT 1200
CATGACTGA AGGAAATTA TTAGACTAT ATGTATCTC TGTGTCTTTT AACTGTATCA 1260
TATCTAGTC TCGATGCA CTTCTTACA TACTGTACAT ACCTGTGAC ACTGTGGA 1320
GTCTGCACT CTTTATCAT CTTGTTTAA CTGTGTGGC ACAAATCTC TTGTCCAAAT 1380
AATTTTAT ATTAGATCT ATAGAGAG ATATATACAC TTTTGAATGT TTCTTAGATG 1440
TCTAACATA AATGCAATTT GTGACTCTA TTAATAAAA TTAATAAAA TCGAGGGGG 1500
CCCGGTAC 1508

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2514 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

GAGGACTCA CACTTCTTTT CATTATCAC TACGATGTA GTGCAATG CAGGGAGAA 60
GCACTTACT GTTGTGTGA GATTGTGTA TGAATCTCAT AACCTAAGAG AGGAATTTAT 120
AGGCTTCTG CTTATGAG CCGTCCAGA AATTTGGCT GTGAATTTTC AGACTATGAT 180
AACTGAGAG TGGGATTA ATATCGAGTA TTGTGTGGC CAGGCTTACA TTGCTCTAG 240
TGAATTTCT TCGAAATGA AGTTGTGTC TTCTAGACTT TTAGAAAT ATCCCGAAGC 300
TATCTACAA CTCTCTCTT CTTGTGCTT AATATCTG TTGGAAAT CAGTACTCT 360
TATGCGATA TGTGTGAT TAGAAACAT TAGGAAGTT TTTTCTTTT TCACTGATC 420
ACCAACTG CTTTACAG TTGCAACT AATTTCTGT CTTTTCAGA ACATTAAGA 480

540 AAGGGGTAAA GAGCTGAGG AATCTGCC TTCTCAGTGG AGAGGAGGG ATGATGCTTT
5 TGAATTTTAA GTGAACTCC TCGAAGACT TGTTTTATGT TTAGATGGTA TAAATAGTGA
660 CACAAATATT AGATGGATA ACTATATAGC TGGCCGAGCA TTGTACTCTT CGAGTGCAGT
720 GTCAATTTT GATTTCTG TTACTATGTT TGTCTTAAA AATGCTCTAT CTTTTACAG
10 AGCCTTTGGG AAAAACCTCC AGGGGAGAC CTCGTATGTC TTCTTTGGGG CCGTATGCTT
840 GACTCGATTA CTCATTTGAC TCAACGAGT GATTCGAAA TATTCAGTTT TATCTGATAT
15 TTGTGTTTGA GAGGCCGA AATTGGGA CCAACTTGA TATTCMAATG AATCTCCCTG
960 GGAATTTCCG CAGAGCTGAC CAGGCTACT TCGATCTCA CTTAAGCTCT GAGCTTACT
1020 ATAAAGAAC CTAAAGTGC CCAAGCTGG AGCAGATAT TCAAGAACTT AAGATATAT
20 TCTCAGACA GCACCTGAA GCTCTTAAAT GCTTATCTCT GTTACCTCTA GTCTGGGAC
1140 AACTCAATTT CATTAGCTG GAGGAGACC ATGCTGAGT GTATAGAGAT GACTTACCA
1200 ATCTGACAC GCTGTGAGCT GAGCTTCAT GTTGAGAGT CAATGGAAA CACAGGGGA
25 AAGATATGA GCTTGGCTCC ACATCTATG AAGCCCTCA CTTGCTGAC ATGAAATTTT
1320 TTCTTAATGT GTATGCTATG CTGAGCTCC TGTATATCT TCTGTGATG AAGTTGAGA
1380 ATGAGCGTAA TGAATATGGA CAAAGCTTC TTAAAGATA TTGAGAGAC ACTTTGACG
30 ACCAAGGTC AAGTAACTG GCTTTGCTTA ACATAAATTT TGAATAAAA CACGAGCTGG
1440 ATTATATGCT GAGCAGATAT ATTAAGCTCT ATACAGTAA GTACAGCTTT CTCAGATTA
35 ATTCCGAAC TGTGGAAAT ACTTAAGAGA CTTTTAAAA TAGGCTTTCT TATATTTGAT
1560 ATTGGAAGA AAAGCGTAA AGCTATGTA GACCACTTAA TCACATAATA TCTTTGGCTA
1620 TAGGACTCCA TTGAATACAT TAGCCATTTA TATCTACTCT GTTTAAATGG CCGCTGTTTG
40 AACTCTGAG CTTTGAAGAC CTACCTGTTT TTGAGAGGA GAGCGTTGAA AGTGCATGT
1740 TTCTTTTTC GTATCTGTC TTGATGGCAC TCTGGATTTG TTTCAGTTAA GTCAATTTAG
45 ACATAGCNTT TATTAAGCT GTGATCTCT ACTGTGCGG TGTATGATAT TCTTTGAGA
1860 AATATATTTT GAAGAGGTGT GAGAGAGAG AATACATTTT ATAAATGTTT GTATGAGAC
1920 CCAATATGTA CTTTGAATA ATAGAGTTT TATGATGTTT AAAATCTAT ACTGAGAT
50 TACAGAAAT TACCGGAGAA AAGCTTGTGA GCTCACAAA CAGGATTTCT AGTGATGAT
2040 TTGCTTTCTT TGAATTAAA GAAGCAATG ACAAATTTG AATGGAAAG CTTGCTGTTG
55 TTCCACATCT GCTGCTGTT TACATGCTT TGTGAGGCT ACATCTCTCT AAGCTTTTAA
2160 GCAAGTATAT GTTGAAGCT TCTGTTTCAAT GATTGAGACA GAATGAGAG CCAATGATAC
2220 TGACAACTGA TTGTGCTGTT TTTTCTCTCT GTCTTTTTC ATGATCTTAA TATATCTGCT
60 2280

5 CACTCTGAT TATAGCAAA ACTCGAAA CCTACAAAT AAGCTTTGTC GTTATCTAG 2340
AAGATATGG AAGATATGTC TGTATTTT GTTGAAGAAA ATCAATTTTC TATATTTAT 2400
TTCAATCTAA ATAAATATG AATTATGTT AAGCTTAGG CACATTAAT TTGTGGGGT 2460
CAAAATCTC TTGTATTAAT TCTTTAAG AATTGTAAG CAGCTTCACA ATTC 2514
4
(2) INFORMATION FOR SEQ ID NO: 256:
15 (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2357 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 256:
CTGCTTATG AAGCGATGC AGAATTTTG GCTGTGAAT TTCACTAT GATPACTGAG 60
25 AAGTGGGAT TAAATATGA GTATTTGCT GGCAGGCTT ACATTTCTC TAGTGGATTT
TCTTCAAAA TGAAGTTGT TGTCTTGA CTTTATAGA AATATGCCA AGCTATCTAC 120
ACATCTGCT CTTCCTGTC CTAAATATG TGTTTGGCA AATCACTAC TTTATGGA 180
30 GTATCTGTC CATTAGGAC AATTGAGA GTTTTCTCT TTTTCCATG ATCCACCA
CTGCTTTAG AACTTGACA GTAAATTTCT GTCTTTTTC AATCAATTA AAGAGGGGT 240
35 AAGAACTGA AGAAATCTG CAAATCTG TGGACAGCA GCGATGTC TTTTAAATTT
420 TTATGAGAC TCTGCAAGC ACTGTTTTA TTTTATATG GTATTAATG TGAACAAAT
480 ATTAGATGA ATAACTATAT AGCTGGCCA GCAATTTGAC TCTGCAATG AGTGTCAAT
540 TTGTATTTA TTCTTACTAT TGTGTTCTT AAAATGTGC TATCTTTTAC AAGAGCTTT
600 GGGAAAGCC TCCAGGGCA AAGCTGTAT GTCTCTTTC GGGCGGTAG CTGACTCA
660 GTACTCAAT CACTCAACA AGTAATGA AATATATGA GTTTATCATG AATTTGTTT
720 TGAAGGCC ACAATTTTG CACCAACT TGAATTTCA ATGAACTCC CTGGAAATTT
780 CCGAGAGCT CACAGGGTA ACTTGATTC TGAAGTACC TCTGAGATTT ACTTAAGA
840 AAGCTTAAT GTCCCAACG TGGAGCAT TATTCAGAA CTTAAGATA TATTTCTGA
900 AAGAGCTTC AAGGCTCTTA AATGTTATC TCTGTACCC TGAATCATG GACACTCA
960 ATTCAATAG TCGAGAGAC ACATGCTGA CATGTATGA AGTGAATTC CCAATCTGA
1020 CAGCTGTCA GCTGAGCTTC AATTGTGAG AATCAATG AAGACAGGG GAAAGATAT
1080 AAGGTTCCG TCCACATCT ATGAGGCTT CAGCTGCTT GACTATAGT TTTTCTTAA
1140

1200 TCTGTATGCA TTGCTGAGG TCGTGTGAT TCTTCTGTG ATGAGGTG AGATGAGCG
GTATGAAAT GAGGAAAG GTCTTAAGG ATATTGAGG AACATTTTGA CAGACGAG
5 GTCAATGAC TTGCTTTTG TTATCATAA TTTCATATA AACACGAC TCGATTAT
GCTGACACA TATATTAAC TCTATACAG TAAGTCAGG CTTCCTACG ATATTTGGA
10 AACTGTGGA AATAGCTAG AGACTTTTA AATAGCTT TCTTATATT GATTTTGA
AGAAAAGC GTATGTGAT GTAGACACT TAATCACTAA ATATCTTTG CTATAGACT
CCATTCATA CATTAACCT TGTATCTTA CTGTTTTAA TGGCCCTGT TTGACCTTC
15 AAGCTTTGA GACTGACTG TTCTTCCAG AGGAGCTT GAAAGTCCA TTTTCTCTT
TCCGTGACT CTGTGATG CACTCTGGA TTGTTTCACT TAAGTCATTT TAAGCATAGC
1740 ATTTATATC ACTGTGACT TCTACTGTT GCGTGTATG AATTCCTTA AGAATATAT
TTTGAGAGG TGTGAGGA AGGATACAT TTTATTAAT GTTCTATGA AGCCCAT
1800 TGACCTTGA CATATAGAG TTTTAAGTAT GTTAATAAT TATACTGAC AGTTAAGA
1860 AATTACCGA GAAAGCTG TGAGCTGAC AACACGAT TTCAATGAG ATTTCCTT
1920 TCTTGACTT AAGAACAA ATGACAAAT TTGAATGAA AAGCTGCTG TTGTTCCA
1980 TCTGCTGCT GTTATATC CTTTTGAGG CTTACATCTT CTTAGCTTT TTACAGTA
2040 TAATGTGAC ACTCTGTTT CATGTGAG ACAGATGAG AGGCCATGA TACTGACAC
2100 TCAATGACT GTTTTATC TGTGCTTTT TCCATGACT TTATATGCT CTTCACTG
2160 ATTTATAGC AATAGCTGA AACCTTCA AATAGTGT GTGTATATC TATGAAATA
2220 TCGAATATAT TCGTATAT TTGCTGAG AATATCAT TTGTATAT TTATTCATC
2280 TAATTAATAT GTATTTTG TTAAAGCTT AGCCATATA TTTTGTGCG GGTCAAMCA
2340 TTCTGTGTA AATCTC
2357

(2) INFORMATION FOR SEQ ID NO: 257:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 689 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

60 ACTTCTGTT GCAAAAGAT GTTCAAGCTT TATTTATAC TTGCTGACC CTTTCCTTT
120 CATTTATGG ATGAGCTGC AGCTATAGA AGAGCTGTTT TTTGATAG AGATAGCAT
180 CAGGAACCG GATGTGGTG CAGGCTGTC TCTGCTGT TCGAGATTG TCGACCTGG

AGCTCTTAGT GGAAGAGCT AAGAGATATG TCGAGTACT TCGATCTTC TCTCTGTTC
240 CGATTTAGG CCGACCCAC AGGTAGCTT CCAATTTTTC TCTCTTTCA TAGCTGTAG
300 GCGCTTCTG GGAATGTTTC TCAATCTCT TAATCTATTA TTGATTCAT TTTCCTCAT
360 GTCCCAAGC TCCATCACT GCGACCACT CCCACAGAG ATGCCCTCT CATTCAGCT
420 GCGCTTGAAC TCCCACTG TGTACCTCT TTGTGTGAG GCGCTGCTC CAAAAGCTTC
480 AGCAAGAGC TTTCAAATG GAGTTTCTA CTGTGAGGS CTTTACATC AGCAACGA
540 AATCTCAT TCTGTGAGG GTCTGCTCT ATTAAATGC AATTAATATG TAAGTACATA
600 AAAAAAGCA TAGAAGAAC GTATGCTTT ATTCTCAAT ATGATGTCT AATAGAAA
660 GCGAATTA TTAAAGTAG TAAGCAAT
689

(2) INFORMATION FOR SEQ ID NO: 258:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2377 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

TGACCCAG COTGCGGA TGTGATGTT CTGCTGATT CCAAGAACG GCGCTATGC 60
120 ATCTCTTCA TATCTTTCAC TGTATAGG GAGCGCCCG GCGCTGTCT ATCTGTGTC
GCGCAAGCTT GCGTTGTTT TGTCTGTAT CTGTGTGCG CCTTGACCT GCGTGTCTC 180
TCTTGAGA AGCTGTATC TATGAACT GCTGACAGC ATCATCTACA GTCAATTCG 240
GCGCTACCTG ATGAAATCT TCCAGACTC GCTGTTTGCG AGCGCGGTG GAGCCCGCT 300
GCTTTGAG TCTATCTC CATGTGGG GAGGAGAG CTTTCCCTCA GCGATTTGG 360
420 GTGAGCCCG AGAATCTCT GAGGTGCTT CAGAGGTTC AGCTGACAG CTGCCACGA
CAGGCTTGA TCGAGAGCT GGTTCCTAT GCGATGTATC TCTCTGAG TCGAGATTT 480
CAGAGCTCT TCAAGAGCT TGACAGAGT GTGTATTAAG AGCAACCCG GAGCGCCGAG 540
TACAGCTC GTTTCTTCA GAGCGACCA GTTCTCTTC GCGCACTAC TACTTTGACT 600
AGCTGGGGA CTTATGCG CTGCCAAC TGTGTGAT TTGCTGTTC CTGCTGTG 660
720 ATGAGATGT TGTGCTGTC TGAAGTGT GACTTCATCC TGGCGGTCT CAAGCTGCTC
780 TTGATGTGT ACTAGCTTT GAGATGCTG GCTGAGATC TTTTGCCTG GCGCTGTGGA
840 TGATCTATK CTTAACCCA CAAGTGT TTGAACCGG TCTCAAGCT TTGTCTGCG

5 TGGAGGAGG GATCTCAACT CTGGCTGTGT ACCGATTGCC AGACCCAGGC TGGAGGCCGG 900
AATGCTGGGG CTGCTGTGTG CTGTGGGACA TGACCCCGCAT ACTGAACTATG CTCATCTGTCT 960
TCCGCTTCTT GGTATATCAT CCGAGTATGA AGCCGATGGC GGTGGTGGCC AGTACCGTCC 1020
TGGGCTGTGT GCAAAACATG CTTGCTTTTG GCGGATCTCT GGTGGTGGTC TACTATCTAT 1080
TTGCAATCAT TGGGATCAAC TTGTTTAGAG GGTGTCTATGT GGTCTTTCTT GGAACAGACA 1140
GCTTGGCCCC TCCCAATAGG TGGGCGCCTT GTGGAGCTTT CAGGCACTG GATTACTGGG 1200
CCAGCACTTT CGATGACTTT GCGGCTGGCC TGTGTCACTCT GTGGCACTTG ATGTGTGTGA 1260
15 AGACTGTGCA GGTGTTTCTG GATGTATATC GCGCTACTTA AGGCGCTGTG TCCAGATCT 1320
ATTGTGTATT GTGTGGCTG GTGTGTCTG TCATCTGGGT CAACTCTTTT CTGGCCCTGA 1380
TTCTGGAGAA CTTCCTTCAC AATGGGACC CCGCGAGCCA CTTCAGGCC CTTCCTGGGA 1440
20 CCGCAGAGGC CACTTACAG ATGACTGTGG AGCTCTCTTT CAGGGATATT CTGGAGAGC 1500
CGGGAGGA TGAAGTACA GAGAGCTGA CCGAGCAACC GACCTCTGTG CTGTGCAAGT 1560
25 GAGTTCGGG TCTGCTTCC CAGCAGGGGC GCGAGGAGAG AGAGGCTGGC ATACACAGG 1620
TCCCATCAT GGAAGAGGG GCAATCTGT GTCCAGCCAG CAGGAGAGAG AGCTTTCTTC 1680
TGACGAGCA CTAGCTGGG GACAGGACC AGTCTCTTTG GTGTGGGCC AACACCAT 1740
30 TACAGAGAG CTGCTGGTGC TTGAGGAGG GCGCTGGCCC TCCGCTTTCT TTATATAGTG 1800
CTTCAATGAG AATTCCTTGG TGAGCTCCAC AGGAGCTTTT CAGACAAA TCCAGAGAGC 1860
35 AGCGGCTCC CTTGTCCCTT GAGCTTTCC TGTGTCTTTT GCTGCCCGCA GCGCTTGGGG 1920
ACGACAGGCC TGACAGGGC CTGACAGGT TTAGCTTAC AGCTTCCAG ACTTCCGGGG CATTCAGCTG 1980
GAAATGATAC TATTAAGTCC GATTTAGCC CAGCAGACA GGTATCTTTC CAGTTTATT 2040
40 TTCTTTCCAT AGCTGTAGG CCGTTTCTG GATGTTTAT CATCTCTCTT ATCTATATT 2100
TGGTCAATT TTCTGTGATG TCCCGAGCTT CCGATCACTG CAGCCCACTC CCGACAGAGA 2160
45 TCCCTGCTC ATCCGACTGG GCTTTTGAAT CCGCACTGT GTACCCCTCT TGTGTGGAG 2220
CGCTGCTCC AAACCTTCA GAAAGAGCT TTCCAAATGG AGTTTGTAC TGTGAGGCC 2280
TTTACATCA GCAAGAGCA ATCTACATG CTGTGAGGG TCGTGGCTCA TTAGATGCA 2340
50 ATATATATCT AGTACATTA AAAAAAAAAA AAAAAA 2377

(2) INFORMATION FOR SEQ ID NO: 259:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1193 base pairs
(B) TYPE: nucleic acid

55

60

- (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

5 TCTGTTCCC GTGCGCCCGC CCGTGGCTTT TGGCGGTTGG GCGCGACTT CTTGTGTGCT 60
ATTTCAGAG ACTCCAAAGC GAGCGCCGGG ACTGAAGGTT TGGTGTGCGA GCGCTCTGGC 120
10 AGAGGTTTAA CTTGGTTCAA ATCCAGCGAT TCTCACTGCG TACGATTTAGC CTGTCCCGCG 180
GCACTGCCG GAGGATTTGA AGTCTCTGAG CCTCAAGTTT GATCTTAGTC CAGAGAGGCG 240
CATGAGGTTG CCGCAACCGG CAGCGCGGAG CTTTCTCTGT AGGCAATTT GCTATTTTCC 300
15 CCGAGTCTTT GCTGCCAGAG CTGTGACTGC GATTTGCGAA GTCTTTCAAG AGCTTCAGAA 360
GCGCTTCCG TACGTCCCGG AGCCTATTA CCGGATCTT GATGGGACC GCTTCGCGGA 420
20 GCTTTTGGC AAGAGACAG TGAACTACTAG TCTGAATGTA TACCGAAATA AAGATCCCTT 480
AAGCAATTT GTATTTGCG GAGCTGTGAC GCGAGTCTT TTTAGCATTA AGCTAGGCGCT 540
GCTTGGTGG TGGTGTGGG CATTAATGGA GCTTTGCTGG GCACTCTCTT AGAGGCGCTG 600
25 CTGATGCGT TTCAAGACTA CTCTGTGTGAG ACTGTGTGCG AAGGAAACA GAAGATGCGA 660
AAGGACTGC ATGAGCTAAA ACTCGAGAG TCGAAGGCA GACTACAGT TACTAGAGAA 720
30 CTGCTGAGA AATTTGAAG TACTTTAGAG GAGATGAGC CTGAGATGA TCTTAGAAA 780
ATTGAGCAC TGTAAAGCT TCTAGAAAC CTTCACTAA TAAATAACA AGACAGGAGC 840
TGAATGCTCT CTGACTTGA ACTCTACTGG AGAGCTGAG GAGCTGCCA TCTCCAGTGA 900
35 ATCCGAGAG AGAGGCACT CTTTGTGAG CTTCTGACA AATTAGTG CTGCTACTG 960
TGTGGCACT GCGTGTCTCT TGTCTTTTTC TTTCTTTTTC ACTTAGAGT GCGGCTGTGG 1020
40 TACTCTCACT TTACTATCC TTAAATTTAA ATACATACTT ATGTTTGTAT TAACTATCA 1080
ATATATGCAT ACATGATAT ATCCAGCCAC CTAGATTTTA AGCTTAAT AAACATTTTC 1140
45 GCAAAAGATT AAGTTGAT TTTTCACTTA AAAAAAAAAA AAAAAA AAA 1193

(2) INFORMATION FOR SEQ ID NO: 260:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1262 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

GAAAGACCA AAGATGCGA CAATCTCTTT GAACATGAT TGGGGGCTCT CAATATGGCT 60

GCATTACTAC GAAAGAGAGA AAGGCAAGT CTCTTAGTA ATCTTGCC ATGTCTTAG 120
GGTTTGTCT TGAACGGGA TTCTGCANT GAAAGGAGC TTGTAAAMA TGAAGAGCC 180
5 ACCATAMAC AAGCTTACAC GAGTSTCCA ATGTAGACA ATGAATTAAT TGTATGAGT 240
CTTGCGTAT TTAAAGGGA GACTACTGAC CATCTCCAG GACATGAAA GACTGAGAT 300
ATAAACTTT CAGCTCCAG TATCCAGAG GAGCTGTGT TGTCTTANG CCGAGTTCA 360
10 ATATGCTAT TTGTGACTG TCTCTTCCA GAGTGCATA TTCTTTTGA AAGTCTTTG 420
GTTTGACAG CATTAAGTAC AAGAGGAA AAGTTTATC ACCGCTGA AAGAGTGA 480
15 GAAATTTGAT CTTTAGAGC ACTAGTTTG GCACTTAA GATTTTACT TAAATTTTAC 540
ATAGTATTTG AACTCATGCA AATATATGT GAAGACATCT AGATTAGTA GTTTATCTG 600
GCGTTTGT TAAACTGAA GATTTTGGAA AATGTTGTC ACTGCTTTC CAGCTATGA 660
20 ATATTTTGT GAATGGAAC CATGATTTA TGTCTGATC ATCCATPAG ACGACAT 720
TTATTTGAAA AAGATGAT TATCAAGAT AATGCTCAC ATTTGCAAT ACTATGTTT 780
25 AAGACCAAG TGAAGGGA TGTGCTGTA GTTGGGTGG TATGTTTAA GCGAATTTT 840
ACGAGAGA ACGAATAA GTTTTGTCT TTGAAGAT TTATTAAT TATTTATG 900
GTCTTTTTT TAAATATAT GTGTGATG TTACATGTA TGTTCATGT CTTTGACCC 960
30 TAAATGCTTT TTGTGATC AAGATGTTG TACTATTTT ATTTTATA ATGTATGTT 1020
GCTTTGCTT GTTTAGAT TACTTGTCT TTGTTATC TTATCTGTA TGAATCTAGA 1080
35 CATTAAGTAT CAGATPAGA TGTTCATGCT TTGAGATAT TACTGCTTG TGTCTATTT 1140
GTTGACAGC TTAAACAGA GTTGATGTA CTTCATAT AACTCATGA TACTTAGGG 1200
CAGTTTCTT GCGATGCG CTTTTGGAA GGAATAAAT TACCCUAG GCAATCCCA 1260
GT 1262

(2) INFORMATION FOR SEQ ID NO: 261:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1179 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

GGCAAACTTT CCCCAGAGC TTGAAACTT GAGAGCGAA ACTTGAATC GTTAAAGTT 60
GGTTGCAAC GCGCCCTCG CCGAGAGG GCGATTTGCG GTTCCGGA GTTTGCCCT 120
55 CAGCGCTCG GAGCGAGCC ATGTCTGCA CCGAGAGC GCGCCCTCG CTACAGAGC 180
CTGAGCTCA TCTTCTGCG CCGAGCTCG CCGAGAGC GAGTTTCA CTGTAGAGA 240
GTTTCTGCT GACTGCTGT TGTACTTCTT ACCCGAGCG GTACAGAGG AGAGATCAC 300
60 ACCATGAG CTACAGAG CTATATGCA GAATATGTT AAGTGTGCA ATGACTCTCA 360

CTGAGCTCA TCTTCTGCG CCGACTCG CCGAGAGCG GAGTTTCA CTGTAGAGA 240
GTTTCTGCT GACTGCTGT TGTACTTCTT ACCCGAGCG GTACAGAG AGAGATCAC 300
5 ACCATGAG CTACAGAG CTATATGCA GAATATGTT AAGTGTGCA ATGACTCTCA 360
CGATGAGT CTATATGCT TGTACAGA CAGTGCAGA ATGTGAGC TGAATTTT 420
10 GAGTCTCT CCGAGCGT TTGAATGAG TGTAGTTCT TTTCATCA AATTAGACT 480
TCTTCTGCT TTTTATGAT GTTTCAGAA CCGATGACT GAGCATTTT ACCCGAGT 540
AATCGGAG AGCTCTATG CCGATTCCA GGAAGCTTT GATCAGCTT GTACAGAT 600
15 CATTCGAG AGGAGCGAG AGGAATCG AGCGAGCG CTCTTACT ACTCGAGCT 660
CTTGTGAG GCTTTAGCG CCGCTCTGA TGAATCAG ACCCTTCA GGTATATGT 720
20 TTTCAGTT TTATGAGT TCTGAGCT TCTGAGCT TGAAGAG CAGAGAAC TGAATCTA 780
TTTGAAGC CACTTTGTG GATTCAGAA CCGAATGAT GAGTATCTA TGAAGCTCA 840
TGAATGTA AATGAGCA CAGTTGCTT GATGAGAT GAGAGAGC AGCTTTAA 900
25 CTTATGAG ATCTGCTTA TCGGCTGTT AGCTACGA ATGTGATC CTATGCTG 960
TATGTGCT TCTATTAC AGCGAGCG CTATGAGC GATGCACT TTAGATTA 1020
30 CTACATGCA CAGTTGAG CAGTATCAC GTTCAGAA CAGACTACT CAGTGTCT 1080
ACCTGAT TATGATGAT TTAATATGT CTTGTGGA AGCATTTCA GAGAGAGC 1140
35 GAGAGAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1179

(2) INFORMATION FOR SEQ ID NO: 262:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1162 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

GGCAAACTTT CCCCAGAGC TTGAAACTT GAGAGCGAA ACTTGAATC GTTAAAGTT 60
GGTTGCAAC GCGCCCTCG CCGAGAGG GCGATTTGCG GTTCCGGA GTTTGCCCT 120
55 CAGCGCTCG GAGCGAGCC ATGTCTGCA CCGAGAGC GCGCCCTCG CTACAGAGC 180
CTGAGCTCA TCTTCTGCG CCGAGCTCG CCGAGAGC GAGTTTCA CTGTAGAGA 240
GTTTCTGCT GACTGCTGT TGTACTTCTT ACCCGAGCG GTACAGAGG AGAGATCAC 300
60 ACCATGAG CTACAGAG CTATATGCA GAATATGTT AAGTGTGCA ATGACTCTCA 360

CCGATGAGT CTTATATGCC TGTCAACAA CAGTGGCAA ANTGTGCAC TGAATTTGT 420
GGATTCCTC CGAGGCGAT TTGATTCAG TGTAGATTCT TTTCAMATCA AATTAGACTC 480
TCTTCTGCT TTTATGAT GTTCAGAA CCAATGACT GAGCATTTTC ACCCCAGAT 540
AATCGGGAG AGGCTCTATG GGAATTCCA GGAAGGCTTT GATCAGCTTT GTACAGAT 600
CAATGCCAC AGAAGCCAG AGGAATCCG AGGGGGAGC CTGCTTAAGT ACTGCCACT 660
CTTGTGAGG GGTTTTAGC CCGCTCTGA TGAATTCAG ACCCTTCAA GGTATATG 720
TTCCAGTTT TTTATGACT TCTAGACT TGGAGAGC CAGAGAAAC TGGATCTTA 780
TTTGCAGAC CACTTTGCG GATTGAGA CCGAAGTAT GATATCTTA TGAACCTTCA 840
TGGATGATA AATGAGACA CAGTGTGCT GATGGACAT GAAGAGAC AGACTTMA 900
CTTATACAC ATGTGCTTA TCGGGTGT AGCTGACCA AATGTCAAT CTATATGCG 960
TAATGTCACT TGTATTAAC AGCAGGCC CTATGTACA GATGCCACT TTAGCATTA 1020
CTACATGCA CAGTTTAC CAGTATTCAC GTTCCAGCA CAGACTTACT CACTTTGCT 1080
ACCTGCCAT TAGATATAT TTAAATGT CTTGTGGGA AGCATTTCA GACAGACAG 1140
GAGAGAAA NANGAAG AG 1162

(2) INFORMATION FOR SEQ ID NO: 263:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 735 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CGGGCTGGT ATTGCTTCG CACCATGCG CCGAGGGCA AGTGGGCGAC GAGGGGAG 60
AAGCAGATAT TTGAGAGAA CAGAGAGACT CTGAAGTTCT ACTTGGGAT CATACTGGG 120
GCCATGCCA TTATGCTCT TGTGAGTTG GTCTCTCTTT ACTCATCTG CTCATTTGG 180
GGCTGGTGG CTTTGGGCTT TAGTGTGCA GTTATATGGG CAGACTTACA CTCATATGAC 240
TGGATGGAC GAGCAGGCTT CTTGTGAGA TGGGGGCTG ATGGATGCTG GACAGGCTC 300
AAGCTGAGC AGGGCATGCG AGACACTT AAGGATGTGA TCTTACTAC AGCATCTG 360
CAGGTGCTCA GCTGCTCTC TCTATATG TGTGCTTCT GGTTCCTGCG TCGAGGCGG 420
GCCCTTACC TCGTGTGGT GAATGTGCT GGGCCCTGGT TCACTGCAGA CAGTGGCAC 480
CCAGACGAG AGCAGATGA GAAGGCGAG CCGGAGCAG AGCGGGGCA GATGAGGCG 540
TTATAGCAAT TGAATGTG GGCAGAGCC ACTGGGCTG GGTGGCTG TCAAGGTGCA 600

CAGCCCTCA TCCCTGAGC AATGAGGTC TAGTCCAGG GCCAAGACA GTCTGAGGTA 660
TTGGATATAC TTATCTCTA TAGGGTGGT GATTAATGCG CTTAGATGT GAAAAAATA 720
AAAAAATA ATTTT 735

(2) INFORMATION FOR SEQ ID NO: 264:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 783 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AAGTGCATGA GCTCCGATG TGTGCTTAG TGAATGGGT TTGCTGGCT CTGCGTGT 60
TCCCGGCTG GGTATTTGCC TGCACCATG GCGCCCAAG GCCAATGCG CACGAGAGG 120
AAGAGCGGA TATTTGAGA GAGCAGAGAG ACTCTGAAT TCTACTGCG GATCATACT 180
GGCCCAATG CCAATTAAG CTTGTGAGG TTGCTCTCT TTTACTCAT TGCCTCTTT 240
TGGCTGCTT TGGCTTGGC TTATATCTG CAGTGTATG GGCAGCTAC CACTCTATGA 300
GCTCATGCG ACAGCAGCG TTCTCTGAG ATGGGGGCT GATGATGCT GCGATGAGC 360
TCAATATGA GAGGGGATG GAGAGATGAG TGTCCGCCAC CCGCAGCCA GGCACCTTAA 420
GATGTGATC CTACTGAG CCAATGTGCA GTTGTGAGC TGTCTCTCT TCTATCTCTG 480
GTCTTCTTG CTTCTGGCTC CAGCCCGGGC CTTTACTCT CTGTGGTGA ATGTGCTGG 540
CCCTGCTTC ACTGCAGCA GTGGCACCC AGCAGCAGAG CACATATGA AACGCGGCG 600
CCGAGAGAG CCGCGGAGA TGAAGCGGT ATAGCCATTG AGCATTTGC SAGTGGCAC 660
TGGCCCTGG TGGCTCTCT AGGTGACCA CCGCCCTGAG CTTGAGCA TGAAGGCTTA 720
GTTCAGGGC CAAAGCAAT CTGAGTAT GGTATTAAT ATACTCTATA GGTGCTTTGA 780
ATA 783

(2) INFORMATION FOR SEQ ID NO: 265:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1638 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

5 GCGACGAGCC GCGCGCAGCG GTGCGCGCGG CGCCCCCGCG CGGAGAGCCCT NCCCTTTCCC
GTCTGGGAGC GCGCGGATCG GGTCCAGCGG ANCCCGGGAC ACAGAGAGCG GGAAGAGAT
GAGTTCCCGG GCGCTCCCGC CGGATCGGAA GAGGAGGAA GTGATCGGA ATCTGCGCT
AGTCTCGC AGAGCGATG TCTACTACTT CAGTCCAGT GTTAGAGT TCAGAGCAA
10 GCGTCATGTC CGAGGTACC TGGGAATAC TGTGATCTC AGCAGTTTG ACTTCAGAC
TGGAGAGTGT ATGCTAGTA ATTACAGAA GACAGAGAG AGACTCGAA AGATCTCT
15 CAATCAAAAT AGCGTAAAC CAGCTTGA TACAGATG CAGTTAGAC AAGAGGATC
AATTTTCAA CAACCGTAA CCAAGTCAC AATCACTCT AGTAATAG TGAATCAGA
CCAGAGCA ATGATGAC AGCGTAA AGCGTCA GCTTTCTCG GAGAGAGCG TACAGAGAT
20 TAGTCAATCA GATGTAGAC AGCAATAT AAAAAAGT GAGTACCA AGGTCTTCA
AGAGTCTGT CAGGTAGCA ATGATGAC CTTTATCT CTTTATCT CTTTCTCA GTCTTTCA
25 CAGAGCTCT GCGCAATCA CAGGCAAGT CTGCGCTCT GTGAGAGAG ACCCTGCT
TTGCTTAC ACATCTAC CCGTCTCA AGCTTTAT AGCTTTAT GTCLAGT AGAGATCAG
GAGAGAGAA GAGCGATAC AGCAATAG CAGAGATG CAGAGATG TATGCGAA
30 CAGTCTGCG CAGGCTGT ATCAGAGA GATGATAT GAGATAT GAGATGCA GTGAGTCA
AGCTTAGAA TATGATAG TAACTTCA CCGCTTTC CAGAGAGAA ATCTTAGAA
35 ATTAGCAA AATTTTCA GTGCTTTG CCGTTTAG CCGTTAGAA AAAATATGA CCGAGGCA
TAGAGTTT TATAGGAT AAGCAATCC TTTTATAG TATTTTGT TATATATCT
ATTATCAA AATCATTT TATTTGAT CTAGAGCT AATATATCT TTTTATTA
40 TAGAGGAA CCGTAGAT AGCTGAGT TTTGATCA GTGCAATCT ACTGAAATG
TAGACTTAC GTAAAGAT TGTTCGCC AGATTTTA TAGAGAGCA TCAGAGATC
45 TAAATAAAT TCCAGTTA AGATATTT GACTTACT TATATAGA TATTTTATA
CTTTATGA AGGAGACC TGTATCT TCCATCTA CTTAGAGC AATATATGA
TATATATCA AGATATG GATTTTTC TGTAGAG CAGCAATCT AAGAGAGCC
50 TGTATGCT TCTGTAT TACTTAC TGTATGCG GCGCTTAG TTTGATG
GAGTGGAG GAGATAGG CATATATG CAGTATGAC TATGCTCT GCGAGTTCT
55 GAGATCTT GCTTACCA GATTTTCA GATTTCTG CTTAATATC ACTTAGCTG
TGTATTTT TTTTCTT

60

(2) INFORMATION FOR SEQ ID NO: 266:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1455 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

10 GTCTGCTACT GCGATGCGAG TACCGGTCC GGAATTCGA GGTTCGAGCC AGCGTCTCC 60
TCAGTTGCGA AGGTACCTCG GAAATACTGT TGAATCTGAG AGTTTGTACT TCAGAACTCG 120
15 AAGATGATG CTGATTAAT TACAGAGAA CAAAGAGAGA CTGCGAAGCG ATCTCTCTAA
TAAATATAG GTTAACCCAG ACTTGATATC AACTTTGCCA ATTAGACAAA CAGCATCAT 240
TTTCAGACAA CGGTAAACCA AGTTCAGAA TCATCTTAGT AATAAATGCA ATTCAGAGCC 300
ACAAGCATG AATGACAGC CAGTCACTT TTTCTGGAG AAGAGGTAC AAGGACTTAG 360
TGCATCATG GTACAGAC AATATATAA AACCATGGA CTACCGAAG GTCTTCAGG 420
25 AGTTGTCCA GTAGCAATG ATGAGAGCT TTTATCTCT GTTGCAGTG CTTTGCACAC
AAGCTCTCG CCATCAGAG GCGAGTCTC GCTGCTCTG GAAAGAGCC CTCTCTTTG 540
GCTTAACACA TCTCAAGCC TCTCAAGAG TTTTATTTTC ACAGATGAG ACATCAGGAA 600
ACAGAGAGG CAGTACAGC AAGTACGCA GAATTTGGA GAGACTGTA TCGCAGACAT 660
CTTGTGCGA GCTGCTGTA CAGAGAGAT GGAATTTGAA ATGAGCAGTG CAGATGAGC 720
CTAGCATAT GATCAGTAA CTTTGCAGCG ACTTTGCCA AGAGAAATTT CTTAGAAAT 780
840 GAGCAAAAT GTTTCAGTG GCTTTTGGT GTAGAGAAA AATGTATCC GAGCAGATG
900 AGCTTTTAA TAGCACTAC CAGTGGCTT TTAGATGAT TTTTATGTA TATATCTAT 960
ATTCAAAAA TCAATTTAT TTTGAGTCT AGCAGTTAA ATTAGTCTT TGTATATCA
1020 AGCAGAGCC TAGATGAG CTGAGCTTT GATCCAGGT GCAATCTACT GGAATGTAG
45 CACTTACGTA AAGATTTGT TTTCCCGCA GTTTTATTA GAGCAGATCA GGAATCTTA 1080
ATAAATTTG CAGTTAGCA TTTTGTGAC TTTCTGAT TATAGATAT ATTAGATAT TTTTATCT 1140
TATTTAGAG GAGACCTGT ACHTCTTCC ATCTCATG ATCTCATG TAAAGACAA TAAATATTA 1200
TATTCAGCA CTGATGGA TTTCTCTGT TGAAGAGC AGCAATTA GAGCCCTCG 1260
TTAGCTTCC TCTGATTTT ATTCAGTCT GATCCGCGG CTTTATGTT GACATGAG 1320
55 GTGCGAGAA GATAGCCAT ATTTTTCAG TATGACTAT TGCCTCTGG AGCTTGTGAG 1380
GAATGTCT TTAGAGAA TTTCTAGGA TTTCTGCTT AATATACG TAGCCTGTG 1440
TAAATTTTT TCCCT

1455

(2) INFORMATION FOR SEQ ID NO: 267:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1086 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

5 GGGCTGCACT ACCGTCGGG AATCCGGG TCGACCGG GTCGCTGAC CAGGAGAG 60
CTGCTGCTCT ACATCAGCTT GGGCTGCGAG CGCTTCGCGC GCGCGGGCGG CGAGCTGAC 120
TATGTGCTCT TCAAGGGCGG CAGCGTGTG CATTCATCTT TGTACCGCCA GCATCTAGCA 180
GTGTGTGGAT GTATGAGCCA CTCAGGAAT GTGTGTGGA TGAAGCATC CTGTGAGAG 240
CAAGCGACT TTATCTTTC CTGACCTTG CTGCTATGAC ACAGCTCTC CTGACTGCA 300
CTGTACTCC TTGAGAGAG AACTCTCTTA GGGAACTTG ATGGGAACA GCGATGCCA 360
AGGACATCTT GGGTGAAGCA GGGCTAGACT TTGATGACT GAGCAGCTG AGGCTGTGG 420
ACCGAGGT TACGAGAG ACATAGAGC TGAAGGAGA GTGGAAGAG TTGTGTGACA 480
AATATGCCA GTTTCAGAA ATAGTGTGG GTTTAATGA GCTGTGTAT CAACTTCCA 540
AAGAGCGA AATGAAAG ATGAAGGCCA TGGTGTGCG GACTTGTCT AATCTATAG 600
CAAGCAGAG AGAAGCTCA CAGCGCAC TTCAAGCCT AATAGCAGA AAGAAATGC 660
AGCTAGAG GTATCGGTT GATATGAG CTTGTGTAA ATGAGAGA GAGCAAAATG 720
AATTTATGA CCAATTAAT TTTCAGAAAT GAATGAAA TTTCGCTTT ATATGAGAA 780
GCGAAACAA AAAAAAGCT CTCAAAACA AAAAAAGCTC TGTAGCATC CAGCGCTTG 840
ACCAATGAC TATGTCAAA GAGTGGCT GTAAAGCATG CAGCGCGCTG AAGCAGCAC 900
TACAGTCTG GGGAGGCCAG TTTTAAATC AGTCGAGC TGTGTCTGTT GCGCTGTGAG 960
TGTAGTCTT CACTCTTAT GCTTAGTGG AACTAAGAG TTGTAAACT TTCACTCTTT 1020
TTTTGTAAA TTCAAGAC TTGGAAGCA GAGCAATTA AATTTGTAT TGAATGCG 1080
TTGATG 1086

(2) INFORMATION FOR SEQ ID NO: 268:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

5 GGCACCGGAG CAGCGGGCT GTCCTGCTG CAGCGCGCG GCGCGGAGTG GCGCGCGGCA 60
GGAACATGCA ACCTTGGAGT TGGCAGCAT GAGTGAATC GAAATACCGG TGTCAATGAC 120
AGCGCGGTA TGTGCTGAC ATATGCAATG GAGTGGCTT TGTCTCATAT TGTCTTACTC 180
AGCATTCCTT TCTTCAGTCT TCTGTGCTT TGGACTTTAA CAAATATAT ACATATCTG 240
GGATGTACG TATTTTCCA TCGAGTGAA GAAACAGCTT TCGAACTCC TGAACAGGCT 300
AAAGCAGAG CTCCTAACTC ATTGGAGCA ACTGAGTAT GAAATACAT TTACATCTTC 360
ACGAGGTTT TTCACAATTT CTCATATAT TCTATATTTT CTGGCAGTT TCTATACGAA 420
GTATATGCA ACTCACTCA TCTTAAGAC AGCTTCTCTC CTGAGTGTAC TAATTCGCA 480
AATGCCACA CTACATGCTG TTGGATCTT TCGATTTAT AGTATATGAA ATGTTTGA 540
ACTGAAAAA AATTTTACAG CTACTGAAT TCTTATAGG AAGGATGCT TAGTAACTG 600
CACTGTTTCT CTGATATCT GAATCAGAA GTATTTACAT TGGAGGCCA ATGCTGCTC 660
CTTCAGTCC TGTTTTGAAG TCCAGATTC CATTAATGA TGGCTCTCTT TAATACACT 720
GTATCATTT TCGAAGGCGG CTTTATAGC AGGCTGGCCA GCGCGAGCTT ATAGTTTAA 780
GCGCTGACA GTGAGGCTCT ATGATATTA TTGAGGAAA TAGAGATTT GTAGAAACT 840
AGGACAGCT TAACTTATA TGAATGCCA TTGCTTTAG AAGAGCAT TTCACTGCTAT 900
TCAGCTGTGG TTATTTAAG CAGACTTACA TGTAAACCGG AATCTCTCT ATACAGTTT 960
ATTAAAGATT ATTTTATTA CCGTAAGAA AAAAAAAA AAA 1003

(2) INFORMATION FOR SEQ ID NO: 269:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1214 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

ATCAGACTCT ACAATGACA TATTTTGAT GTGTGTGTG TGTACTTCA AAGTAACTAG 60
GAAAAATTA TCTTCGAC ACAGTACCT TGTCAATGCA GATTTGCGGG TGTAAAGTTG 120
CCAGTTGAT CAGTGTGAT TCAATTCAT ACTTCTACA GAGCAAACT GAACTTTGCA 180
GTTGCCACA CTGAATGAA TCGAATACC GTTGTGATG AAGCGCGCGG TATGTGCTG 240
ACATATGAT TGGAGTTGG CTGTCTCAT ATGTCTTAC TGAATATCC CTTCCTCAT 300

5 GTTCTGCTTG CTTCGACTTT AACAATATTT ATACATATTC TGGGATATTA CGTATTTTTC 360
CATCGATGCA AAGGACACCC TTTCGAACTT CCTGACCCAG GTTAAGCCAG GCTGCTTACT 420
CATTTGGAAC AACTGGACTA TCGATGACAG TTTCATCTTT CACGGAAATT TTTCATCAAT 480
TCTCTCATTA TTCTATTTTT TCTGGAAAT TTCTATACGA AGTATGATCC AACTGACTTC 540
10 ATCTTAACA CAGCTTCTCT CCTGATGCTA CTAAATCCCA AATGTCOACA ACTACATGTT 600
GTTCGATCT TTGGATTTTA TATGATTTGA AATGTTTGA AACTGAAAA AATTTTACA 660
GCTACTGAT TTCTATTAAG GAGGATGAG TTATGAACT GCACTGTTTC TGTGATATG 720
TGAATGACA AGTATTTACA TTGGAGGCC AATGCTGCTT CCTTCAAGTG CTGTTTTCGA 780
GTTCAGATTT CCATTAATG ATGCTCTCTT TTATACACC TGTACATTT CTGAGAGGG 840
20 GCTTTATAG CAGCTGGCC AGGCCAGCT TATATGTTAA AGGCATCAC AGTGAGGTTG 900
TAGTAGATTA ATTCAAGGAA ATAGAGATTT TGTAAAGAAC TAGGACCCAG TTACTTATA 960
ATGAATGGCC ATCTGTTTA GAAAGACCA TTTCCTATCA TTCACTCTTG GTTATTTAAA 1020
25 GCGACTTAC ATGTAAACC GATCTCTC TATPACATTT TATTAAGAT TATTTTAT 1080
ACCTACATA TTTCCTCTTG TTATATAG YGGATGATA TCTCTTGT TTATACAGC 1140
30 CAGTTCCAC TATATAGGAT ACTTTTTCG TTCTCTGCG CTATATATG TGTATGCTC 1200
AATGAGCCA TTTTACAT TATATAGTT ACAG 1234

(2) INFORMATION FOR SEQ ID NO: 270:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 574 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

60 NGAAGTCCCT TGTAGCCCT CTGTCTCG CCAGATGCT TCAAGTATTT ATTAAMACA 60
TATGATCCC CATTAGCCC TATACCCA AATTTTACA GGAGTTTGG ATAGGATGG 120
GGCTGATGG GTTATGCTTT TATTAATCC GGGCTGCTGA TAAAGAGAT AAGCTTTGA 180
AAGCTTACG GCTGCTCTT GGTATGCTT AACGATTTT ACTTGAGTA CATGTGAAG 240
55 AATAGCTAG TGTGCTGTA AATTTACCA AGCGGTGTTA GATGGGAGC GTGAGAGTC 300
ACTGTACAT TGTATAGTA CGGTACTT CATGGATTA ATAGGATAT ATAGGATAT 360
CACTGTACC TGTACTGCT TTGATGCT TCCCTCCAG CCGTCCGGG TGTACGCAAT 420

ACTCTGAGTA GATATTTTGT CATCGAGCC ATGCATACAG ATCTCACTG AGCCACCAAT 480
CATTTGAAA TATTTACTTC AGTTGTACAG GACTTGCTCA TCGAGATCA GGCATCACT 540
5 TGTATTTTAC TGTCAATATA ACUTTTATTA AACT 574

(2) INFORMATION FOR SEQ ID NO: 271:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1731 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

20 GCTCCAGGTT GCGCTCTGTC CCGCTGCACA TCCAGCTCAC TACCTCGGA AATCTTACAC 60
CTTCAGCAC TGTGTTTTTC TCGTCTGATA TCGAGGAAG GTTCAGACCA GGCATCAAGT 120
25 ATTTTGGGGA TATATATAC GTGGACACA GATTTTCCA AGGGGCCCCG ATTTTAGGAA 180
TTCTGTTTAT TGTACAGAA CAATACCTTA AAGTCTCTCG GAGCAGGTTT CAGAAATTC 240
ATTTACAGG TGTAAACTG GTACTTCCA AGACCAAGTT TTCAATGCTA TTACAGAG 300
TAGAGCCGC ATTAGCAGAG ATTCCCGAG TCAAGAGTGT TGTATATTT GAGATGAAA 360
CTCATGTCG CATTACAAA ACTGCCCTCG AGCTATGTC CCGAGGAGTC GAGTTTACA 420
35 TTGTTCTGA TCCCACTCA TCAAGACCA TGAATGACAG GATTTTGGC CTCGAGGTC 480
TGCCTCAC CAGGATCAT ATGACCCG AGTGAGGCT GTTCTCTTC AGCTGTAC 540
TCAATAGGAC CATCCAAAT TCAAGAAAT TCAAGATCTA ATTAAGGCA GTGCTCAGA 600
GTCCGCTCTG CTTTCCAAAG TATAGGACAT TTGAGAACT GGTATGCTAC TCACTGTGA 660
720 AAGACAGTCA GGTGAAGGAC TGTAAAGCCA CACAGCTCT TCTTATCTCT ACTAGATTA 780
45 AATGTTAAG TCAAAAAGG CTCCTTTTTT GGGCTCTCTA GTGACTTAA CCACTAGAC 840
CAITTCAGTA CCAGCATTA GTTACAAAG TCAAGGCTT CCGTCTCTG TTACTTCTT 900
TTTTTGTAA TGTGCTTTTA TTTATTAATA AATATTAATA TGAAGATGCC TGTTTTGTCT 960
50 CTACTGTGA CTCTGATCT ATCTTTCCA AGTCAGACT CTGTGAGCT TTCTTAAAT 1020
TGTCACTTT AAGAAATG AGTACCAAC AATGATTCG GTTTATAT ACTGTAGAT 1080
55 GTTATATGT TATGTGAT GTAGTCTTT TACTTTACAG ATTGATGGA ATAGATATAT 1080
TCAATGAAA TTTACCCA GAGCTCTGAA TCAATTTAC CACTCCCTC ACAATTTGT 1140
CCACTATG AGTTGATG ATCTATCCCT ACCAATGAT GTTCAATAT TACTATCTT 1200
60 TCTGACAT ACTGATTTCT TATTTGTC ACTATTAATA AATCTCTCT AATATTTCT 1260

CTTTTAAGTC AAGAGGATG GATAGAGAG GTTTCAGATG CCAATATATTT GGTGAGAGGC 1320
TGTATTACCA TCTTTGAAGT ATGGCTTGGT GATATATCTTT ACCAGCATCT TGAATATATA 1380
TTCTAGTGTG CACAAGATTT AGCAAAAGA TAAAGCTTGG GTGGAATATC ATTATTAAT 1440
GTTCACTGTC TCTTCTATAT TTCTCTACG TACTCTCCAA ATATGTATAT GCAAAAGTGC 1500
TCAGTAAGCA TTCTGATGTA TTAAATTTGT GGTCAATGTT TCTCTTCGAT AAATTTATTT 1560
TCATTAATA CTTTTPAGAG GGTTTTGAAA TGTTTTTCAA ATATGTGAAA TGTGAGACTG 1620
CTGTCTTTTA TATTAAGATA ATTAAAGAA ATGTATGTG ATTGAATTTA TTTTGGCTCT 1680
CACAAGATG CTCATGAGT ATCTCTCCAG GAACTCTAT ATTATTTAA G 1731

(2) INFORMATION FOR SEQ ID NO: 272:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1320 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

CTCTTAGGA AGAAGAGTGC AGATTTGCG GGGCAGAGC GATCTTGGC GGTGCGCCAG 60
TCAATATGA GTGAGGGGC AGACACCTTC CCGCAATTC TGAAGGTTT TTAGTCTGGA 120
CTAGGCGAGT AGCCGAGAG TCTTAGTGGC CGGCTTCAG TCACTGCGCG CTGAGCGGAG 180
CTGCGCTGCG CATGTTTGG TCGTTTGTGG CGGGAGGCT GTTGGAACA GCTGCAAGGC 240
AAGTGGCAGA GATTAATTT GTTTTGACT TACCTGATTA TGAAGATATC AACCATGTTG 300
TGGTTTTTAT GCTGGGACCA ATCCATTTTC CTGAGGGGAT GGGAGGATCT GTCTACTTTT 360
CTTATCTGTA TTCAATGGA ATCCAGTAT GGAATCTCT AGGATTTGTC ACGAATGCGA 420
AGCCAGTGC CATCTTCAAA ATTTCAAGTC TTAAATCTGG AGAAGGAGC CACATCTTT 480
TTGAGGCAAT GAATATTTTC GAACTCGAT CTGTTGCTCA GATTCGAAT TCACTGGAAT 540
TAATAGAGAG TATGCTCTGG CAGACTCTGG TAGGTATGCG TCTGTATCC TCACTTGAAT 600
CAATCACTCA GTTCACAGA AAGATGTTGG ACAATTTCTA CAATTTTCT TCACTATTTG 660
CTGTCTCTCA GCGCGAGTGC AGACAGGCC CATCTGAAT GTTCATTCGG GGAATGTTGG 720
TTCTGCAAT GTTATGAGC AATTTCTGTC TCAATATTA AGGCTTTTAA TAACTGATTA 780
TCAATTTGT CTAATGAAT ATTCCTTTT TGAATTTAA ACAATTTCT TTATTTGAAA 840
CATCAAGCT GATCCGCAAT AAGATATGTA CTATAGAGAT CTGATGAGA ACAGTTCTTA 900

60

CCCTAATAT TTTTATATAT TGTGGCAAT ATGAATTTAT AAGAGAGCA AATATATTT 960
GCTTATGTTT TAGGAGCAC TATTAAGCT TATTAATAT TGTATATTT CATTTAGAG 1020
TACCATCTAT GAGATGATTT TATACTGAC TGTGTACATG AATGCTAAT GAATCTATTT 1080
TCCAACTTTC CCGTGTTTTA TAAATATTT TTTCACTTT GATATCTTA GAGATGAGAG 1140
GATGCTAGG AAGAGTTTGT TGAAGATGT TACCATGTTG TAGCATGGA GAGCATTTGG 1200
AATGCTAGG GTTGTGAATTT GGCATATGAG TACATATGTC ACCTGAGCA AATTTCTCTC 1260
ATCTGCTCAT CTGAGGATG AGCAATAGG ATGGAATTTG AATTTCTCTA GATCTCTCTA 1320

(2) INFORMATION FOR SEQ ID NO: 273:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

CCCTGAGAG GGGCTGCTGT GCGAGCTTGG GAGAGTCTG GATGCGGCT GCGCTGATG 60
GCGCTGATGT GGAATGCTT GCGACGATCT GCTGGGGTGA ACTTATTTT AGCCCTTCC 120
TTGTGTTCT TATTAAGAAC AGAGAGGGGG TGGCAGATC ATGATGTTCA GCAATGATTA 180
TTCCGAGCAC AGCGCTCTG GAGAGGCAAT GAGCATTTTC TTTCAGAAA TGTCTATAT 240
TCAAGCAGAA GGCATCATTT AATTAATCTC TCACTTTGTC CCGAGCTCTG TTTTATAGCC 300
CCTTGGCCAG ACTGAGAGG GCGAAGGAG GCTAGATTTT AGTTAAGAG GAGCTCTGABA 360
GATATATGG TCCAGAGCA CCGCGGGGGG GTGAAGGCC TGTGAACTTA TGTCTAGGG 420
AGCCGAGGT CACACTTTGA CTTTCTTACC ATGGCTCTG TCTATGAGAG TATATATCT 480
GCGTAATAT TACAGAGCA GTCAATGTC ATTT 515

(2) INFORMATION FOR SEQ ID NO: 274:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

TCAAGCAAT AAGGATTTCA TGAAGAAAT AGAAGAAAG CCAATGAGAG TTAACTAGC 60

AGTACGACATG GAGAAATGTA TCTTGCGCTGT CTGCAACCC GATTGCATCA CAAACATCAA 120
ACATGATAC ACCATATCC GAGCTGCGCTT CAGAGAGATC CTGACATGCG CTAAAGACGA 180
5 CAGAGCGGCT CTGAAACGCG CATTCTTGAA ACTGCTGCGT AATCTGAGC TCTTGGAAGA 240
ACTTCTGCGA TGGATCCAGT GGGCTGAGAC CAGCTCATTT CAGCGGATC AGGAGCGAAT 300
10 CCGCAGAC ATTGACCGAT TTAAAGCGCT TATGCTGAG CATGCGATG ATATGAGGGA 360
GATGACTGCG AATAGCGCTG ACTGAGACCG GCTACCGAG ACTATCAAAA GGAATACAT 420
AGAGCTTACT CAGCGCGCTT TCAATGAGAA ATCCCGCAG GAGGCGAGAA AATCCCTTAG 480
15 TCAAGCACCG CTTCTCTCCA TGGCATCTCT TTCCAGATCT GAGCAAAA ACCCAGGAT 540
CAGCAGCTTT TCTGCGCGCT GCGAGCAGGT GTGGCTGTTA GCACTGAGC GGCMAAGGAA 600
ACTGAATCAT GCGTTGCGAT GCGTGAGGA GTTGGAAGAA TTTCGCACT TTGACTTTGA 660
20 TGTCTGAGCG AAAAAATATA TCGCTTGAT GATTCAGAA AATCTCGAG TGAATGATTT 720
CTTCGCGCGC ATTGATAGCG ACCAGATGCG GAGATTAACA CTTGAGGAT TTATCATGCG 780
25 CATTTTAGCA TCGAATGTC CACACAGAA GTTAGAGATG ACTGCTGTG CTGACATTTT 840
CGACCGAGAT GCGGATCGTT ACATGATTTA TTATGATTTT GTGGCTGCTC TTCTATCCAA 900
30 CAGAGATGCG TATGAGCAA CAGCGATGCG AGATAAATTC GAGATGAGC TTACAGAGCA 960
AGTGGCTAG TCGAATGCG CAATAAGGTT TCAAGTGGAG CAGATCGAG AGAATAAATA 1020
CGCGTTCTTC CTGGCGATTC AGTTTGAGGA TTCTCAGAG TTGCGGCTG TCGGTATTTT 1080
35 GCGCAACGAT GATGCTTCCG GTTGTGAGAG GATGATGCG CTTGATGAA TTTTATAGTA 1140
AATATGATCC CTGCGAGCA CAGATGAGAA CTAAATTTGA ACTTAGAGAG AATTTGATCC 1200
TACGAGAGCG AGCATCCGAG GGAATGACCC CATTGCGCTC AGGGGCTGCA AGCTCCAAAC 1260
40 CATCTTCCCG GCGAGCTTCC CTTACTGCTT CAGCTGCGAG TCGTAGTCTAG AGTAAACGCA 1320
GCTGTACATC CATGCCATCT TCTCGAGCCA CCCCAGCGAG TCGAACGAG GTTATCCCAT 1380
45 CATCAGGTAG CATTTGAAA CAGCAGACAC CAACTTTTCA TTCTAGTCTCG ACATCCCTTG 1440
CTGCTGATAC CAGCAATTAG TTCTTCCCGG GCTTCCGAG GTGCCAATC TAAATCGGCA 1500
50 GACCTTAAA AATCTGCGAG TGGCGCTGCG AGTCGCGCTG GAGTCTGAGC CCGGAGTCCA 1560
GCCAGAGAG GCGAGGAG TGAAGCTTCT GACTTTGACC TCTTAGAGC GCATGCTTG 1620
TTTCGACATCT TCGAAGGCA GCGCTGAGCG GCGCAGAGC AACTTCAGCA CAGGCTTAAA 1680
55 CAAGCTTCC AATATCCAA CATTCTGTAA GAGGACGAC ACTGCTCCG CAGAGCTCC 1740
AGTGTCCAG GATTAACAT GTTAGAGAC CCCCAGGCA CTATCACTT TGAATCTGCG 1800
TGCATACAT GGTGTATAT TTATTTGAA CCGGAGAGT TATATGTTA AATGTTAAA 1860
60

AGAAATATG TGTATGAG CTGCTTATTT TTTTTCCTTT TTGTAGTTA CTATTTCTCT 1920
GTGAATATTT ATGTAGTAA AATTTCCTTC CTGTAAACCC TGTAAATGAT GGGGCCGAGA 1980
5 AATCAATAT TTGAGAAA CAGTGTAAA GGTCAAGATA CAATGTGTA TTAAMAAAA 2040
AAAGGCGAT TAAATGCGTT TCTGCGCGT GCAAGGTGTT AAAGCTGCTT TATCTTTTAA 2100
10 GATTATCTCT AATATCATCT TCTTTATAAA CTGACTTGC TATCTGCA AGATTAATTA 2160
TATTAAMAA ATAAAGATCC TCGATGTTT AAGGAATCT TTTTGTGAA ATCAAGGACA 2220
CCTCAATTAG CAGAGCTGA GGGAGGCGT TTTTGCATG TTAAATGTT TGTATTTT 2280
15 AGCTAAGAG AGGGAACCT ATCTAAGTAA CATTTGACA TGGATACAG AAAGGAGTT 2340
CATTCATA CTCTCTTTC ATATGTTTC ACTACTGCT GTTTAAAGCA CAATAGCTC 2400
CTAAGATCA GGGTAAATG TAAAGTTCA GAAAGCTCA GAAATTTG GZTTTAAAC 2460
20 TGAATTTG CTGCTATTC AGCTAGACA CCAATACATC TTGTGTTTCA CAGGACCGAG 2520
ACCTTTGCA AGGATAGCG TCGTTGCTGA CATGTGAT TTCAATTTG TTTTATCCAC 2580
25 TTTTTCCTCT ATTATTTAA ATGCTGATC AACTTCCAC AACTGAGC CTGCTCTTT AATGGAATTC 2640
CGAGCTGTT CTGAATAT GCACTAGAA CAACAGCTG CTGCTCTTT AATGGAATTC 2700
ACCAAGAC TTGTAAACA GTCTGTTTG CTTCCTCTCT TTTTGTGCG TAAATAGTTC 2760
30 AACTGAGCA GTGACATCA AAAGGGGCTG TCTGGGGCTC CTGTTTTTTA CGTGCTGTC 2820
TTCAAGCTCG ACCATGTC TGTGTGATTA TCTCAATTC TTTTAAATTA GCGCAAACT 2880
35 GAGCTCTAC CATGAGCTG TTAGAAACA AGACACACT TTGTATTTA AATGCTTCA 2940
GTAAACAAA AAAAAAAA AAAAAACTCG AGGGGGGCGC GGTAC 2995

(2) INFORMATION FOR SEQ ID NO: 275:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1990 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

GGAAGCCGCG GGGCTCCCGG GATGCTGAG CAGGGCGCTG CTGCMGCTG TCTGCGCTCA 60
ACCGCAGAG ATGAGCTGCG TCGTGGGAT GCGCTTGTCTG GCGTACGTG CTTCTGTTTG 120
GGGCACTTC GTTATATCA GZTCTATCCA GGAATATGCT GACTAATAA TTGAAAGCA 180
GATTGAGAG ATGCTGAC CACTAGAGCA GAAATACGA GATTAGAAA AAAGCTTTAC 240
60 CCGAATATC CCAACATTA AGTTTATATC AGAATAGAT CCGAAGAGA TTTTGATTA 300

5 CAGAGGCGC AGGTTCTGTC TAACTGACA AACTCATGAT GAGCGCCAC 360
GAGGTGACCG TGGTGGACAA TTCTCTACG GCGAGGAGA GAATCGTGGG GACTGTGATC 420
GGCATGAGA ACTTCGAGTT GATTAACGAC GACGTGTGGG AGCCCTCTCA CATTGAGGTT 480
GACGAGTAT ACCATCTGGC ATCTCGAGCC TCCCTCTCAA ACTACATGTA TAACTCTATC 540
10 AGCAGTTAA AGCCGATPAC GATTGGGACA TTAAACATGT TGGGCTGGC AAAACGAGTC 600
GGTGGCGGTC TGTCTCTGGC CTCGACATCG GAGGTGTATG GAGATCTCTA AGTCCAGCCT 660
CAAGTGGAG ATTACTGGGG CAGCTGTAT CCAATPAGAC CTGGGGCTGT CTACGATGAA 720
15 GGCAGAGTG TTGAGAGAC CATGTGCTAT GCTTACATCA AGCAGGAGCG GTGGAGTG 780
CGAGTGGCA GAATCTTCAA CACCTTTGGG CAGCGCTGTC ACATGAGCA TGGGCGAGTA 840
GTGCGAACT TCATCTGCA GCGCTCGAG GGGGAGCCAC TCAGGTATTA GCGATCGGG 900
20 TCTGAGACA GGGCTTCCA GTACGTGAG CATCTAGTGA ATGGCTGCT GCTCTCTGTC 960
AAGCGACCG TCAGAGGCC GGTCAAGCTG GCGAACCG AGAGACAC ACATCTAGAA 1020
25 TTCTCTGCT TAAITAAAAA CTTTGTGCT AGCGAGTG AATTCAGTT TCCTCTGCAA 1080
GCCAGGTC ACCGACGAA AGAAATCCA GACATCAAAA AAGCAAGCT CATCTGGGG 1140
30 TGGAGCGCG TGGTGGCTT GAGGAGGT TTAAAGGAG CAATTCAGTA CTTCGTAAA 1200
GAATCGAGT ACGAGGCAA TAACTAGTAC ATCCGAAAC CAAGGCTGC CAGATTAAG 1260
AAGAGGCGA CTGGCAGAC CTGACTCTCT CACTTTTGG AGCAGAGCT ACCATGTAC 1320
35 ACTTCATGG ATGTATTTT GCTTTTITT TTGTTGCTT TAAAGAAAG CTTTAAAGG 1380
TGTCAATGAG AAGAACTGG AATTCATGTC TGAAGCTTGC TTAAATGAA TGGATGTGC 1440
40 TAAAGCTGC CTCGAAAAA CTCGAGATT TGCCTTGCAC TTTTTGAATC TCCTTTTAA 1500
TGTAAATAG GTTATGACA TCCTCTGCTA TTTCAGATT TTTCATCTT GCTGTGAG 1560
45 CATATGTTT GACTGTGTT GACAGTTTA TTACTGCTT TCCTTTGTA GCTGAAAAAG 1620
AAGTTAAGC GGGACNAAA ATGCGGATT TATTATAAA AGTGGTACT TAAATATGA 1680
GTCTTTATAC TATGCTAAA GAAAAATCT ACGAGTATG TCAGGTGGTG GTGGCGGCG 1740
50 ATTGATTTA GGGCAGTAA AGAGATCTG TGTGAGGCT TATGTCTCT CTTTATATC 1800
AGAGTTTTT CAGGTCTTAC TTGTGAGTG CAAACTTGC TTGAAATAT TCCTGTGCT 1860
CATGATCAG GATTAATGAA ATGACTACTG TGTTTGCTG GTTATCTGG GCGGGGCGAG 1920
55 GTTGGGGGG ACAAAGTAA CATATCTCT GTTAAACATG GTTAAATATG CTATTTTAT 1980
AATAATTTGA 2000

(2) INFORMATION FOR SEQ ID NO: 276:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2436 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

AACTTCGCTT AGCTCTCGAG GGTAAAGCG GTGAGGCTT AAAAAGAGA GAGAAGAGA 60
15 TTTAAAGTCC GTTGCATGGA AATTAAGAAA CATATGAT GTTTTATCA AGCATCTCT
CGCATTCCT CTTCTTTATA AGAGCACT ATCACTATCC TGGAAACCTG TACAAAGGTT 120
TCAGATGGG CAAAGAGAG CAGTGAAGA TACAACTTCA GTTTCACAC CCAAGAAATC 180
TTACAGAGA CAAAAGAG ATGCGAGCA GATTATTAAC CTTCCCACTG CCAATATATG 240
300 CAGCAATTTG GCGACTTTTA ATTATGACA GAGAGAGCC TTGAGGGGAA GTAGAGGTGG
25 CCGAGGTTGG GGCACAGAG GAATCTGTA TCGGGGAGA CTCTACTGAA TAAAGATCA
GCATCTTCA CAGTGTGAT GAGCTTAATA TACTTAATTT CTACTACTCA TTGATTTGCC 480
GGGATATCCC CTTTAAAGAG ACTGCTGCT TCAGCTAAA ACTTAATGTT CTTTATCT 540
30 TTGTATGAT GACCTACTTT TGTAAAGAG CATGTTGTC TCGAGGTAA AACCAAGTGT 600
AATATTTGG ATGCTTTGTC TCGAATCTTG ACTGTGTTT GCAATATCAT TATTCAGACT 660
35 TCAATTTGG AATCTTTTAA ACATCTGAT AATTTGTTGT TGAAGCTGT TCATCTTAAA 720
ATGTAAAGAA ATTCAAGTCA GTTCTCTGTA TAAAGATCAT CAGTTTGTAA AGGTTACTGA 780
TTTCTGCTT CCGCTGTGAT TTTTAAACA ATATATGAG AGAGTATATG GTCAATCTTA 840
40 ACATTTGTT TTAATTTGTT AATAAGCTG CTGGGCACTG GTCCAGATTT CTTACTGAT 900
GTCAATAAG CAAATTAATT ACATAGCTT CTTAAATAT AGGATATGCA TTACTATTTT 960
45 AGGAGAAAT AGTTCCTTT GCAACCCCTA CTTAATCTT TTCCATATAT TGTGATGAA 1020
ACTTTTGAT ATGCAATCTT ACTAATGAA TAAAGATGTC TATGTATAT ATATGATAT 1080
50 ACATAAGAT ATATGTGTT GTGTGTGTT ATATATATAT ATATGATAT TGTGATGAT 1140
GACTACAGA CATTAATCAC TTTTAAATTT CAGAGAGCG GTAGCTGCAC ACGTGTATTA 1200
TCTTTTGGG GCTGAATCGG TTAATTAATT GTTATTTAGG TTTTACTTCC CAGTACGAG 1260
55 GATTTCTAG TTAGTTCGAC TTACTGAT ATGCTTATTT AAAGCTAGA ATAAAGGCTG 1320
CATTTTCAA GATTAATGCG AATGCTGTT GTTCAATTA CACCAAAAT ACTGATCTG 1380
ATGTACTAC AGGTTCTTAC AGGAGAGAT GATTAATTT ACATTTGGA GATTATGAA 1440

504GGGCTAC CCGAAGAG TGACTGTATA AGTGTGACC ATATAGTAG GATCTCTC 1500
TGGTTTGGT TTTCGACTT TCGAGTTT ACTTCTCAG GTTATTATC AATATTATG 1560
5 TATATGTTAG CCAATAGAT TTATTAGTTA AACACAGA TGGGGGTTT GTGAGTTT 1620
TAATCATG GCAATTTTA GTACAGAGA CCGTTTGTTC TGCATTTGAA TGTTTGTAT 1680
ATTATTGTT CACAGTTAT CTTCCTGCC CAGTTTGGT ATTCAATCA ACTGCTGA 1740
10 TCAATTTCT ACTAGTCTGA TGTATTATTC TCGAGATTAG TTGTGATTC CAATCAGGT 1800
GTCTTCATTA CCAATCTCT TACTCTCAG AAGACGAA ACTGCTTAT TGAATTAAT 1860
15 GCAATGTGT ATGGGGAAA TAGTCTCTGA AGCTTAGAT CATACAGTC AGCAATGTT 1920
GTCTAGCTTG GCTATGGAT GTTGCGATA ATCTCTAAC ATTCCAAAG ACCATGCT 1980
20 GAACTTAAC TCCCTTGGAA TCTGACAAA GGAATATAAA ATTGCCATT GAAACTGAC 2040
CAGCTAATCT GCACTCTGA GATGATCAG CCAATGGCC AAGGCAATT CAGTACAGA 2100
AATATAGAG ACTACAGCTA ATAAATTTG AACTAATAT ATAAATTTAC CACTTTTGT 2160
25 CTTTATAGC ATATTGTAA ACTCAGACT GAGCAGAGT GACTTTACTT TCTCAGTTT 2220
GATAGATGT TCACTGTTC CTATGCTTC ACCCTTCCC TTCCCTTTTC TAGGCAATTA 2280
GTCCAGACT TAGGTATTT TTGCTTGGG ATTGATGTA AATATTAAT GCAATGAT 2340
30 TTTTCTTTT GCAAGACCC TGTATATAT CTGTTTAAA TGTAAATGTC CCGTATGCT 2400
TTTGAAATTA ATTGCTTTT GTAAATTTAA AAAAA 2436
35
(2) INFORMATION FOR SEQ ID NO: 277:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 782 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 277:
40 GCACTAGCT TTCTCCACC TTCTGTCTCC CCAATATAG TTTATTTGTT TGTCTGAC 60
50 TCACTCTGG CTTTATTTA ATTTCTTAG GCGCTGAG AAGCATTTT TACTCAGAG 120
GATTAGAGC ACTTGACAA GCGCCGACA TCCAACTCT GCGATTTGTC GTGGAGGAG 180
65 GCACTCTGG GCGATAGAC CAGCAGAT AAGAGAGCT CAGTGTGAA GCGAGAGTC 240
TGCAATGTT ATGATGCTA AATGCGTTA TATCCTTTC CCGTTTAAAT CAGATCTTG 300
TGAAATGGA AATCAGCA AGAGGGGAT CAAGATAGC TGATCTGACA TGTCTCCAG 360
60 CAGAGGCGA GGTGGAGTC AATCCCGGT GACAGCTGG TCGAGAGCCC TGTTTGAGT 420

TGTGCTGAT CCGTCTCTG TATTAGTTT TTCCCTGGGA CAGAGAGCC CTAGAGAG 480
5 GGAATGACAG GGTCCGCGG GATCTTTTC TCCCTCCCT CATTAGGCA GAGCAGCT 540
GCTTGCAC CCGTCTCTC AAGGAATGC CTTCGCGAG AATGCCAC ACATATACC 600
TCTCTTTT TTTAGTCAA ACTCTGTTT ATTCCTTGGC TTGCTCTCT CTTTCTCTCC 660
10 CTCTCACT TTAATCTGA TTCTATTTC ATGGAATTC GATTTGAGT TAAATCAA 720
CAATGCCCC AACACAGT CTTCAGGAA AATATCAA AATATTTAA CAAAAAAA 780
15 AA 782
(2) INFORMATION FOR SEQ ID NO: 278:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 961 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 278:
20 GAGTTCCGC TCGAGAGCC TGTCTGGGC GGGGCTTC ACCATGGCT CCGCAGAGT 60
30 GCACTACCC ATCGAGATC CCGATAGCC CTCTGGAGC CAGAGAGCA GCGCCAGCC 120
AGTGTGAG GAGCAGAAA CTGCGAGCC TGTGTGAT TTTTGGCT GGGGTGGCTG 180
35 CAGAGCAG AACCTTCCA AGTACATGC CATCTACAC AATAGGGCT GCATCTAAT 240
CGATACACA GCGCCGTGC ACATGCTTT CTTCCTGAG TCACTGGGTA TCCCTTCACT 300
TCTGTTTTG GCGCAGAGC TGTCTGAGCT GTCTTTTGT TATGAGTTG AGAGAGCC 360
40 CCGCTCTTC CATGTCTTA GAGAGGTGG CTTCAATCTG TACGCTTAG TGTCTGAGCT 420
CTCAGAGC CCGCTTCT GCGGCTGCG TGTGTGGGC ACCATCTTG ACAGGCTCC 480
45 TGTGTACAG AACCTGTAG GGGTCTGCG GCGCTTCCA GCAATCTCG AGCGCGGCG 540
CGCATGCTG CCGCTTTC TGTCTGGC CTTTGGCTG GTGTCTCTC TGTTCAGCT 600
50 CCGCTTCT CCAATCAG CCGCTTTCCA CACCACTTC TATGACAGC TACAGAGCC 660
GGCTCTGCG TCGCGGAGC TGTACTCTTA TCGAGGGCT GACGAGTAG TCTTGGCAG 720
AAGATAGAA CCAATGTGG AAGCAAGCT GCAAGCGCG GTCTCTGCG GTTCTGTGAA 780
55 TTCTGTCTA TCTGACAGC TCGAGAGCT CCGTACTAC CTTACTTACT ACAGAGCT 840
CTGTGTGAG TTGATGCGA ACTGCTGCG CTCTGAGC CATCTCTCA TCTCAATCT 900
60 GCTCAGAA TAAATGCTG AATCTGCC ACATATGCG AAAAAAAA AATATGAGG 960

6

961

5

(2) INFORMATION FOR SEQ ID NO: 279:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1228 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

CGCGCTTTTC AGTTGGTGTCT CTTGGTGTAC GGCCAGGCC AAGTAGGGGA TTGGCTTCC 60
TCAGTGGCA GCGCTATCAG ATTGGATAT GTCTTCAATA TTTCATTGGA TTTCACATGG 120
TTTCACAGAT GTCTACAGT TTTTAGGAT ATATAGAAA ACTGTAAAC TGTATTTCT 180
TGAATTTGAT AATCCAGAA AATCAATAT GCTACATAT CTAAAGATG ACAGACTTGG 240
ACAGACTGTC CGACATATAC ATCCACTTC CGAGAACTG ACCTTCTGTC GCATGAGTT 300
TACACTTTT GATCTGGGTC GAAATTTTCA AGCTCGAGAC CACGAAAGGC TTTTAGAGTC 360
TGTATCAAT GCGATGTAT TTTCTGTGA TTGTGCGAGC CACGAAAGGC TTTTAGAGTC 420
AAGAGAGAA CTGTATTCAC TAATGACAGA TGAAACATTT GCTAATGTGC CTATACTGAT 480
TCTTGGGAT AAGATGACA GACTGAGGC CATCACTGAA GAGAGCTTGC GAGAGATTT 540
TGTATTATAT GTTCAGAAA CAGAAAGGG GAGTATATCT CTGAAGAAC TGAATGCGGG 600
ACCTTAGAA GTTTTCAAT GTAGTGTGCT CAAGAGACA GTTTAGCGAG AAGGCTTCCG 660
CTGATTTGCA GATATCATG ATTAACAAA ACTACATTC GTTCCAGGTC TCACAGTTTCA 720
GCTTACTCA GAGATTTGAT TGTCAACAT GCTAATCTTG AATTCATAG ACTTTTCTG 780
GTATTAAGAC AAGATTTTAT TACATTAATA ATATTAAATC AACTTAATTT GATATGATAT 840
TGAAAACTGA TTCAATAG TTGTAGTATC AATATTTAG ACATGTTAG CTTCCTAATT CCATTAAGT 900
ACTTGTTTT PACATTTTAT AATGTGACAT CACCCGAGG CCKTTTGTAA AGAGCACTT 960
TCCAGCKATA CATTGAGGC ACTTTTAC ACATGAAAC TATTAACCAT ATTTAAGGC 1020
TCAATATAT AATTTTATA TGTACTTTTC TGGAACTAGT TTTTAATTT PAGATATAT 1080
GTCCACCAT CTKATGTATA CATTATATA TTAGCTTAT CANTGATTC ATGATGCTT 1140
ACAGTTTCA ATACTTTTTT TTCTATGCA AAGTCAATC AATTAAGAA ACTCTAATAT 1200
TTGCAAAA AAAAAAAA AATTTGCA 1228

60

(2) INFORMATION FOR SEQ ID NO: 281:

(2) INFORMATION FOR SEQ ID NO: 280:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

TTCTGGGTCT CCGGACAGGT GAGCAAGCTG ATGAGGCCA CGTGTCTTAT GCGGCAGCTG 60
GCGGTGGCA GCAATGCTG GCGGCTCTCC CGAGGCGGCG CGGAGAGCGG TTACAGGTGA 120
TTTCTGATTT TTACATGACC TTGAGCAGGT TTGCATATTA TCGAAGGCA TCGCTTTCTT 180
CTTACATAT TCTGTATAT ACACATATCA TCAGTGAAGA GTTGTGAAA GAGCTCAGAG 240
GCTCTTCA CCACTATPAC CGAATTTAGA TCGACCAACA CCGGAGCTTC AAGGAGAAC 300
TACTCATAT GTGTGAATGG TGGACCAAG CCGACATCTC CTTATGTGAG CAGAGATTC 360
AAGATTTC AATAGCGGAG GTGTGTAGAG AATCCATATC AATGCTCAGG GAGGATATTA 420
AGAGCTTCTT CAGCACTCT TACCATACA AGTTTCCCTT TTTCTATCTT TCTTCCGCGA 480
TTGTGATAT CCTGAGAAA ATATTCGAC AATGAAATGT GTTCCAGGCC AAGATCCACA 540
TGTGTCTTAA CTACATGAT TTTAATGAG ATGCTTTTCT CAGAGATTT AAGGCGCAC 600
TATTAACAC ATPACAGAG AAGAGCTCTG TGTGTGAAA CTGTGTATC TTCCAGCAAC 660
TTGAGGCCAA AACCAATGTC ATCTGCTGG CAGACTCTAT CCGGAGCTTC ACCATGCGCG 720
ATGCGGTTC TGTGTGTGAG AACTTTTCA AATTTGCTT CTTGAATGAC AAGGTGAGAG 780
AAGGCGCGAA NCGCTAATAT GCACTCTTAT GACATGTGTC TCGAGAGCA CAGAGCTCTG 840
GATGTGTCA ACCGCTACT CGAGCAATC CTGTGCGAG GGGGTCCAG TCGAATATCA 900
AAGGCGCGCA AAGGCGAGGC TCGAAGGCC SCTGAGGCC GTGTGAGCA GCGGCGCTTC 960
CGAGATCTT GCTTCCCTCT GAGCAGAG CAGAGCGAG GTGTGAGCA GTGCTGTGGT 1020
CTTTCGCGC CCGTCCGTC TCTTTTCCCT GAGCAGCTTC ATCACAGAG GCTTGAAGCA 1080
ACCGGCAAT GTGTGAGGC AAGGCGACTG TTCTGTGTGA ACTTTGAGC ACGATATTC 1140
AGTGTCTTAG GCAATGTCTA CTCCAGGAT TTTCTTCAAA ATTTTATTAAC ATCGAAGTT 1200
GAAAGATTA TATGTGTGCA AAGAGATCA AATTTTAAA ATGAAGAAA AGCTGCTCTG 1260
ATTGAGGCA TGTGTGTGCG GTTAGAAGCT GAGCTCTCTG GCTTGGGGCG ACATGCGTC 1320
CTTCTAG 1327

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 799 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

10 TCACTCTGCC TACAGCTGAG AGCTGAGATG ACTGCGCCCT CAGCGGTAC TGTGACGAG 60
TGTATATAC AGCTGAGT AGCTGAGG CAGCGCTGCG GTGTGTCCTC GTCACTGTGT 120
GCTTTTGCT GAGCTGAT TCCAGCTG CTGTGTTTTT GAGAGAGCT GAGAGAGCA 180
TAATCTCTGC CTGCTGACC CAGCTGTGGC CTGAGTCTC TGTGAATACA TTATCTTGG 240
ATGTTGGGTT ATTCAGCCCA AGACATTTTC AGTGCCTCT AGCTGATTTG TACATATTA 300
TAAATATCTA TTGAGAAATT GTTCCATATA TGCAGTGTCT TTGCTCTGGG TACAGCGCA 360
GCCCTTCAC CCGACTTGG ACTTGAGGAC CTAGCTGAGT GAGCTTTTCC AGCTGTCTCT 420
AGAGAGGAT TCTGATATCT AGCTGATAC GAGATGTTTT TCACCAAGGT CACAGAGCA 480
TTGCTGCTCT GATGGGTTG AGTTTGGTTT TGGTCTCTCT TTGAGCCCAA TATCTAGAGA 540
ACATTTGAAA CAGTCTGCAC CTTTGATAGG GTATTCGATT TCCAGAGCA CCAATCCATT 600
TTGTGGAATT TATGTCTCTG TGCCTATATA ATCAATGATA CAGCATATAT ACCTTTTCT 660
CCATTTTCTC TCGAGGAAC ATACCTTAAG TTTTTTTTGT TTGTGTTTTG TTTTTTTTGT 720
TTTGTGTTTC CTATTATGAG AAAAATATAA ATAGTCAAT TTTTATATCT AAAAATATGA 780
CAAAAAAGT CAGAGGGGG 799

(2) INFORMATION FOR SEQ ID NO: 282:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

50 AAGACTCTTA ACATCCATGA GCTTGACATF GAGCAAGAGC CTACTGTGTC CAGCGAGATG 60
GCTGAGGCTT TCGTACCTT CCGAGATGGA TGGCTCTCT ACTACAGCA GCTGTGTTTT 120
CTGCTGCGCA TGGGCTCTGC TTTCCTTTAT ATGACTGTGC TGGCTTTTGA CTGCATCACC 180
ACAGGTATAG CTACACTCA GAGACTGAGT GCTTCCATCC TCAGTATTTT GATGGAGCA 240
TACATATAA CTGCAATAT GCGACTGTFA GCTTTTACTT GCTTACCTCG AAAATGTGCT 300

TTGTTTGGCA CAGCTCTGAT CTGAGGATTC GCGAGCTTTT CTTGTTGAT CTTGTGTGTG 360
ATCTCTGAT TCAATGCTGG AAGCGCCCTG GACTTGTGCG TTCTCTCTTT TCGAGATATC 420
5 CAGTCAAGGT TCAITCAAGG AGATGCAATF ACAGCTACCA AGATACCTGA AATTCAACT 480
GAATATACA TGTCTAATGG GTCTAATCTCT GCTAATATTC TCCGAGAGAC AAGTCTCTGA 540
TCTGTGCGCA TATCTCTCTT CAGTCTGCTG TTTCGAGGCG TCATTTGCTC TTAGATCGGT 600
CTTTGTCTCT TTGATTTTAC TGTGACAGG TTGCTGCGAG AAAATGTATF TCAATCTGAA 660
AGAGGCAATA TAAATGCTGT ACAGAACTCC ATCAACTATC TTCTTGATCT TCTGCATTTG 720
15 ATCATGCTCA TCTGTGCTCC AAATCTCTGA GCTTTTGGCT TGTCTGTATF GATTTCACTC 780
TCTTTTGTGG CAGTGGGCCA CATATGTATF TTCCGATTTG CCGAAATATC TCTGGGAAC 840
AAGCTCTTTC CTTCGCTGCC TGAATCGAAA GAGTTTGA GAAGAAATCA AGCAATATCA 900
TCTGTGTTT GAGCAAGTTT AACTGTGCTC ATCTGTTTAC TAGATATATF AGAGCAATG 960
TGTATTTT GTACTGCGA ATTCGATATA ATGCTGGGT GTTTTGTCTC GTTTTATCA 1020
25 CAGCTGTGCC TTGAGAACTA AAGCTGTGTT AGAAACCTA ATCAGCGAGA AATTATCTGA 1080
TTAATTTCCC TTATGTTGAG GCTATCGAAA AAATATGAAA AGAATTAAT CAGTTTAAAT 1140
ACGAGACTA TATATGATAC ACTGAATGCC CTAATTTCTC ATGAGTATAT AGAATCTTAC 1200
30 GTAAAGAGT GATTAGTCA CTTGATTCAG TTATCATTTG ACAGATCTCT ATCTGTACTA 1260
GATTCAGAT ATGTCAGTTT TCTCGAAAC TCACTCTCT TCAAGACTAG CTATTTATF 1320
35 TTTTTCATC TTATTTATTT TTAATATCAA ATTCTTCAG TATGAGACT AAATTTGAT 1380
AACTAATATF ATCTTATTC ATCTATATGA TCTTAAGSTA TTTATATGTA TGTGGAAAA 1440
CAAAAGCTT AACTAGATF CTCTAATGAG GTTATGCTT TACTTTAAG AGCACTTTG 1500
40 TATTTTATF ATCAGATGCG GCAATATF GTATGAGCA TATGTAGCNC TTCAAGCCT 1560
GTTATCATG TAACTGCGAG GTAGAGCAA AGCTGTAAAG TAAATTTATC ACAGATGAC 1620
45 TCGATACAGA CTTCGAATAT GTCAATATGT TGTGATAGA ACCTAGAGC CAAAGGCCNC 1680
ACAGAGGCC AGAATGCCA ATTTAATCTA TGTATCATC ATTAGTATC TGTGTGTAG 1740
AAGATGAGG GTTAGGCTT CAGCTGCGCA AGTATCATGT AGAAGGCCA CACTTTGAAA 1800
50 GCTTTGTTT TACAAATGAC TTGATTTTAC ACATCTAGGT AGATATTTT TATTTTACT 1860
GTTTTATACC CAGAGTTAT TTCTACATTC TTCTACAGCA AGAATATCA TAAAGATATC 1920
55 CCTTTCAAT GCTTTTGA GAATATGAG AAAAATGTT TGTATATAT TTAATTAATF 1980
GTTTAAAG TCAATTTCA ACATGTCTCT ACCAGATG TACTTTGCT TAACTCTTAA 2040
TATGCTTTT CAGTGAAT CCAATGCTT GCTATGAGCA CTTTCTTTT CCTTGAATF 2100

510

511

TAACTCTTGG CTTCATCTTT CTACAGTATG ACATATGAT TTGCTATGTT GTAAATCTT 2150
TGTAAATAAT TTCTATATTA AATATTTGAA ACTTAA 2196

5

(2) INFORMATION FOR SEQ ID NO: 283:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1185 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

10

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

GGAGTTAAGG CTTCGTATTA GGAAGAGAG TCTGACAGA GCACACAT CTGAGCTCC 60
AGAGTGGGG GATGACAGAT CAGATTCAT CTGATTTTC TGTAAATA CTATTACTC 120
ATATGATC TGAACAAAGA TCTGTATTTT ACTGTGGCT CCACTCTCC TCTGGGTGA 180
GTAGATGTTA AGCTGAGCT TGGACGCT CTTACATGA AGAGATCTAG CTAGACAGC 240
AGACTCCCC ATTTATGGA AGAGATTC AATTATCT CTATTTATTA AAGATTTTT 300
TAAATGCTCT TGGTATTA AATCTAATG TCTGGGTGT GATCATCTAG GAGCATTA 360
CTATCTACT TGGATCTT TGTCTAGG GAGATCTTT GGGGGCTGG AGTCTCTCT 420
GTCCAGCT AAAGCAAG CTTCACAG GTAGAGCCA CAGAACCTC GCACAGAG 480
GCGGTCTAG GAGATGAT GTTACAGGA GAAAGGAG GAAAGGGGT GCACAGAG 540
TAGAAGCCA GGAAGGATG CCGACTGA GACCATGG GACACTTA TTGTGAGA 600
GGAGGATCT TCTTCTTGA ATCTGACA CAGCTAGCT GACCTTCT TCGAATGTC 660
AGCTGTTTC CATTGATAG GGCACCTCT CCTGCAAG CAGCAATGT GCTTCTATC 720
AGAGAGAA ATATTCATC ATGTGACA GAGTCACT TCGACTTGC ATGACTCTT 780
TCCAGCCA CAAGAATCC TGGTAGAT GAGATCCC TAGTCTTAG GCTGTCTTT 840
TATAGAGTA CGATTCAGC ACTGCTCTA GTCTCTGAA TACAAAGC ATTTGATCT 900
CATCTGCA TATGAGCTG TTGAGTTTT TCGAGCTAA GTTCATGAC CAGACATGA 960
GCGAGATCT GAGGCGAG TACAGCTG TCGATGAT GTTCTTTTG AACTATGCA 1020
ATAGAGGTT GACAGGCCA GCTTGAAGA CGAAGTGT GAAGAGATG AGGAATTA 1080
TATGATGTT GAGAGGAT CCAATCTCT GATCTACTG AGGAGAGGY TTCTAGTTC 1140
TTTGATGAG AGCTGTTTC CTGCTCTTA GCGAGCAC TCCA 1185

60

(2) INFORMATION FOR SEQ ID NO: 284:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1634 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

AGCGAAGGCG GAGGTAGCG GAGGTAGCG AGGTAGCTC CTAGGCTGAG AGGTATTAGC 60
AGGAGCTGCG TCGATGCCCC TGTATCGAG ACAAACGAC GATCTCTAGT GATGCTAGC 120
AAGCTTGTTT TGTACAAAG CAGGTGGGA GTCTATTTT GTACATGAGA TACATCAGAC 180
TTACTCTGCG GCGATATG TGAATGAGT CTGATGTTT TACATCTATG CTTTCCCTGC 240
CCACACAAA TTGTATCAT AGTCTTCAGA TGAAGACAC CTTTTCGCA GCTGCCAACC 300
AAGAGCTGCT TCTAGGCTCG TGTATATGT CATATTTAGC STTTTATAT ATGACTTTTG 360
ATTCTCTTGG TTCTATTTT AGCAGATGT ATGCAGCTC ATTAAATAT ATCTGTGTGC 420
ATACAGATAC GATATATAT GTTGTGTAT GATATATCT CTATATCTTA GTTTCAGAA 480
GTTCACTGTA AGCAGATGTA GTCTCGAGC CAGGAGACA CCGTCATCC CTGCTATTAG 540
TGTTTGCGAC AAGTATATAT GATCTCTCT TATTAATAT TTCTATTTAG AAGACTGAG 600
CAAAAGCTAT AGTCTTCTCT GTTCTTTTGG CAGCTAAGT AGGATCTTGG GATGACTTGC 660
TGTGTCTCTC AAGCTCAGT TTGGGGGCAAT CTCTGAGTA TTAGGCCCC TTGTCTTGG 720
TGTACTATCG TCTGTGCTCG TGTGTGTGT TGAATGATC TCTTGCATCG CTTCATATGC 780
TGTTTTGTGG CATTTGGGGA TAGGTGCTCG ACGCCAGAGC ATTTGCAAT TTATTGAGC 840
CTGTGTCGA ATGATATATC ACTCTGTGG AGCTGTATG TCTGCTTGGT TGTGCTTTT 900
CCTGCTTTTC TCTTGAAGA GAAAGAGAT CTGTGAGGC CAGGCTGAG TGAATGAGC 960
TGCAGCTGCG TCAATGCTTT CTTAGAGCAG AAGAGAGAT ATGCAATTT ACTAAGTTCC 1020
TAACAAACA TTATGCGAG CAGACTCTT TGCATGTCG GAACTGAGG CACATATAGG 1080
TTATGACTTG CTCAGATATA TGTAGCTCT AGGGGTAAA TCAAGGCATC ACAATTTCTG 1140
TTGAGGCGCG AGAATAGCG TGTGATTTGC TACGACTTTT TTTTATTAG CATATCTTT 1200
TTGACTTTT CCTCTGAAG TCGAAGAGC GTACACTTTT CCCAAATGTA GACTAGATC 1260
TGGAGATGC CAGCACTGT ATAGTTCTGC TTTCGAGAG AGGAGAGAT TTTAGAAGCC 1320
AATGATCTT AATTTTATT GCTCACTCTT GCTTTTTCG GTTAGAATAT TCAAGTATG 1380
AATATTTT ATGAGAGAT GCTTGAAGC ATGGGTATC AGGAAGGCT ACCTAAGTTC 1440
ACATATCTGC AACCAGACA GCTACCAAGC ATTACTTAGC AGCAGGAAA TGAATTTAT 1500

5 TGGATTTCTCTG TGTGTCTCAA ACTAGGAC CATTGTCTTT GAATACATGC CACTCTAAGG 1560
CTGGGGCGGG TGGCTCAC CTGTATGCC AGCATTTTGG GAGGCGGAG GCGCGGGGG 1620
1634
10 (2) INFORMATION FOR SEQ ID NO: 285:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1795 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 285:
20 TTCCCCCCCAG GTTGGCTTCC TTGGATTCTT TTCTTGTGTA TCAGGTTTG ATTGAGAGA 60
CAACCCCTTC TTGTCTAAC TCATATATCA GCTCACTGTG GAGGAGCAG TGGGGCAG 120
CTTCCCTTTA TGTTCATTGT TACCCCCCAA GACCGCAAAA ACTCTGTGTG GACACAGAT 180
GGACCCCTAG CCAGATCTCT CGACAGCTT GTGTCTCTGG CAGCTGAGC CTTGCCCATG 240
TTAGAGAGAG AGCTCATGTA TCCCCGGGGA CTTGGGGACA TCAGGACAGT GTTCTGGGGG 300
CCCTTGGACA TTATAGAGCT GTATATGGC CTGTATCTTC GGCATATCCC GCGGCACCGC 360
AGGCTTTGTG ACTGGCAGT TGGCTCTCTTC TGGCGGGGGCC TCTCTAGCA GCGCGGGGGC 420
TCATATCTGA TGGCGCTGCT GGTATGAT CTATCTAGC TCTATCTGAC GAGCTGAG 480
GAGGCTTTTG GGGATCTGGC CTTTCTTCTC TATGACAGC ATGTGTGAGA GGTATTTGTT 540
GTCTCTGGA AGCCACACAG CTTTCCAGCG CAGCCCTTCA AGGCTCTCAG CACAAAGGGG 600
CGCATGTGTA TGTCTGAGG TGGGGAGCTA GTATATGTGC CCATATGTGA AGCATTTCTG 660
GAGGACTTTG CTGTGTCTGG TGAAGGCTG GTGCAGACTG TGGAGGCGCG AAGTGAAGG 720
TGGACTTGTGT GATCCAGCT CTGGAGCAG CTGTAGAGCG ACGAGAGAC ATTGTACCTC 780
TAGAGCAGTA TGTCTATAG ATGACTTCCA CCGCTCTTGG ACATATATCC TCCATGAGG 840
GCTGTCTGCG TGAATATCT GATCATCTC CAACAAAC CAGCCCGAAC TTCTCTCTG 900
ATGCTCTCAG ATTGGGCGG GGGCATGTTG GCGCATTTAG TCTCTCTGGC CTCACCTCC 960
CAGAGAGAGA GTGGAGCA GCTGAGAA GGAATGAA CAGAGATC CATCCACCTA 1020
TTAGCCCTGG GCTGTGACT CCGTGGATT TCCCACTCTT TCTTATGCT TCTTCCAGAA 1080
ACAGAGAGG GATATGTGAC CTGGAGAGG CTCCTCTCTCC TCTCTCTGTC CAGGACTGT 1140
GCTTAGACTT AGCATGCCCT TCACTGAGT GTCAAGGCTT TAGATGGGAC CACCGGAAA 1200
60 TGTGGCCCTT CTGATGCA TCACCGAGC TGAAGCATGG AAGGGGCTA TATGTGTATG 1260

1320
1380
1440
1500
1560
1620
1680
1740
1795
5
10
15
20
25
30
35
40
45
50
55
60
ATTAGACAC ATTGAGGAG CAGAAATCCC TCTGTGTGTT ATGCCACTTC CAGAGGTGGA
GACATGCGAA AAGACCGAG GACAGGAGAG GATTGGGTAG GTTGAAGGGT CAGGGGACTG
GTATCAGCCC AATCTTGAG AGGTCCAAA AGCACTGGGG GCTACCCGTT AGCTCATCTT
GCGCTGGCTG TTTGCCGTT CATGTCAAAA ACTGCCACTA CTATGTACTT GCAGTGGGGT
TGCAGAGATG GGGGAGATC AAGTCTTACT CCCCAGAGC TCCCAGGGCC CAGAGGAG
AATCTGGCT CTTTCTAGTC TGGTCTAGAC CCACCTTTCTG GTAGGCTCTC TGTCTCTCT
AATCTGGCT GTTTTTCAG ACTAGCTCA AATAGTGGCC CTCCTTAGAG CCATCCCTCG
CCCCAGGCT CAGGTGATCT TTTCTCTCTC TGAATATTTA GAGCATTTAC TGTCTGTCTA
GTCTCTTTGG CAGGACACAA CAGTGGCATA AATCTATTTG TTTTGAATCT TGATT
(2) INFORMATION FOR SEQ ID NO: 286:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 858 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 286:
TCTCTTTTGG GTCTCTCTGG TACTCTCTGG CCGCTTCTTC GCGCTCTCTG GGTTCGCA 60
GCTCTCTGAG GAGATCTGGC CTCAGTTTTC GCGAGCGATG AATCCCTCTT TGTCTCATTT 120
TCTCTGAGTC TTTCCGCTGA AGGTATTTGG CTACAGGCA GATCCCTCTA ACTACCAAT 180
AGCTCTGGCC TTTCTGGAAC TGTCTGCTGG GTTCTCTCTG GTCATGGGCC CAGCGATGCT 240
GCGAGAGATC AGTAACTTCT TCTTATCTCT CCGAGCGATT CCGCTCTCTG GGTTCCTGCT 300
AGCTCTGAAA GATCTACTTA GCACTGTAT CCGAGCGATT GTCTCTCTGG GGTTCCTGCT 360
GCTCTGATGT GTGGCGAGC TTTTACGCA GACTTAGAG GTGGTCAAG CCACTAGGAA 420
GAGACTCTTA AGTACATCA AGAATCTCTG GAATAGAGC ATCTCTCTCT CTTTATGCA 480
TGCAGCTCTC AGCAGGAAA CATGTAGAA CACAGAGTCT ATCATCTTCT TACAGTATA 540
ATATCAGGG TCAGCAGTG TTAGAGAGA CATTTTGTCT ACCTGGCACT GCTTCTCTTT 600
TTTACGTTTA CTACTCTTT GTAGGAGTA CATGTATATG ATATTAACT TCTCTATGTC 660
ATATGAATAT ACAATTAG CAGAAAGAA ATTTAAATCA ACCAAATTC TGATGCCCA 720
AATACCACT TTTATGCCCT TGGTGTAGT ATACCTCTGA ACTTTTCTCT GTGCTTTTAA 780
ACAGATAT ATTTTCTTTT AATGAAATA AAGCATATA TCTATTTTAA TTTCTCTCTT 840

TTAAATACCTT ATAACTA

858

5

(2) INFORMATION FOR SEQ ID NO: 287:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 915 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

15 GAAATTCGGCA CGAGCGCGG CATGGCGCTC CTGCTTTTCG TCTCGCTGT ACTCTGCGC 60
GGCTCTTTCG GCTCTGTGGG GTTGGCCAG CTCGGGAG AGATCTGGC TCCAGTTTCG 120
GAGCGGATGA ATGCGCTGTT GTGACGTTT CTGAGGTGT TCCGCTGAA GTATTTGCG 180
TACGAGCCAG ATCGCTGAA CTACCAATA GCTGTGGCT TTCTGCACT GCTGCTGGG 240
TTGCTCTGCG TCAATGGCGG ACGATGCTG CAGAGATCA GTAACTTGT TTGATTTCTG 300
CTCATGATG GGGTATCTT CACTTGCCA GCTCTGAAG ATGACTAAG CACTGTATC 360
CGAGGATG TGTGCTGGG GTTCTCTG CTGCTGATG TCGGCGCT CTTAGCGCG 420
ACTAAGAAG TGTGAGAGC CATTAGGAG AAGACTCTAA GTACATCAA GGAATCTCG 480
AAGTAGACA TCTCTGTC TTATGCGAT GCACTGTCA CAGCGGAC ATGTAGAAC 540
ACAGATCTA TCACTGTAT ACCAGTATA TATCGAGGT CAGCAGTGT TGAAGAGAC 600
ATTTCTCTA CTGCGAGT CTTCCTCT TTACCTTTAC TACTCTTTG TCAGAGTAC 660
ATGTTATGCA TATTAACAT CTGATGTCA TATGAATA CAAATTAAG AAAAAAGAA 720
TTTAAATCAA CCAAAATCT GATGCCCAA ATACCACTT TTAATGCTT GGTATAGTA 780
TACCTCTGAA CTCTTTCTG TGGCTTAA CAGATATATA TTTTITTTA ATGAATATA 840
AACATATAT CCAATTAT TTCTCTCTT TAAACCTTA TAACTATATA TAAAAAATA 900
AAAAAAAA CTGCA 915

50

(2) INFORMATION FOR SEQ ID NO: 288:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1517 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

55

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

60

CTTCTGTGCA ACTATGTGGT CCCCCGGGT GCAGAAATTC GCGCATGTGT TCTGATCTG 60
AAGATATCTT GATTTCTCTT GAGATATCCA AAGCTTCGGG GAGGAGACC CAAACCCCCA 120
AAGGAGGCC TGAAGAGAC AGGAGAGAC CTCAGGTGAC CAGGGGTAGC CTTCAAGACA 180
CGAGCTTAG CCTCAGAGG ACGAGGACA GCTCTGAGC GCTGAGGCG AGCGGATGTA 240
CGACTCACT TGAAGCTTTC CCAAGAGAC AGTCTGACA GACAAAGTGT GCGGACATGT 300
GAGCTGAGG TCTTCTGCA CAGGAGACA CAGGGGATTT GTCTATGAG CTGACGCCAC 360
CTCAACCTC ACCTGTGACT CAGGACACA GAAGCAAG TTCTCACTCA AACTGATGTC 420
CAGGATGCG GCTTGTCTA ATGAGCAGAA CTTCCTTCAG GGGGCGCCA AGCTCTGCA 480
AGTCAAGAG TGAAGAGAC TGTACTGAC CCACTGCTG GCAATCCCTA CTTGCAATGG 540
TTTCTGTGTT CACAGAGCA AATAGAGTT CTTCGTGTTA CCAAGCTTGG GAGAGAGCT 600
TCACTGCGC CTGATGTCA GCGCAAGCA TGTGCTGTGC AAGAGATCT CTGCTCAGG 660
TGGCTTCGG GCTGCTGAT GCTCTGAGT TCTTCAATGA GAATGATAT GTTCAATGAA 720
ATGTGACAG TGAATATAT TTTGTGATC CAGAGACCA GATTCAGTGT ACTTTCGAG 780
GCTATGCTT GCGTTTCGC TATGCGCAA GTTGGCAACA GGTGCGCTAC GTGGAAGGCA 840
GAGAGACTT TCAAGAGGG GACTTGAAT TTCAATGACA TGGACCTGCA CAGGATGTC 900
GGGCGCTGCG GCGCGGTGA GCTCAGAGC CTGGGTATAT GCAATCTGAA GTGTCTTAT 960
GGTATCTGCG CATGACAAA TTGCTTTCCA AATATGAGA CATCATGAG CAANAGACA 1020
AGTTGCTGCG GAATCATAT TATGTAAAG TGAATCTTGT CATGCCAAA ACAAAGCTGT 1080
GTCAACCTCA CAGTTTTCGA GTTTTCACC TCCGCCAGT TCTTCCGTGT GATTGACCA 1140
GATATCTCG TTATGACGC GCTTCGCGGG GACCACTCC CTGCTTTTGA GTCAAGCACA 1200
GACAGCTTAC TTGAGCGGCC GCTGCGCCC CACATTCAC TGAATCTGTC GATGCCACA 1260
GTGACGCTT CTCAGGACA GATGACCTC CTGAAGTGA GCTTCTTTC TTGGAACCTA 1320
CGAGTTAAA TCTGCTCAA ATGTTTGAT ACCGCCATT GCGCCCTCAC AGCCAGAGC 1380
TCCCTGACA GTTGTGCTGT GTGTGTGCT GTGTGTGCT GTGTGTGCTT GGAACGGTGT 1440
GGGAGCTAC CTTTGGTGT GCGGTGTGCG CCGAGAGCT GTAAATATA AATCTTAT 1500
TTCCAAAAA AAAAAA 1517

(2) INFORMATION FOR SEQ ID NO: 289:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3865 base pairs

(B) TYPE: nucleic acid

55

60

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

5 TCGAGGGGGG GAGCTTCTCT GAGAGGTGGG CCGAGGCTTG GCGCTTCGCA CTTCATCTTC
TGAGCTTTCT CTCTCTCAT TTGGTGTGAT GTCTTTCTTG CAGCTGCTTT TCAGCAGG
10 TGGTTCCACT GGGGGGAGCT AGCGTGAAT GACAGGATG GGAAGCGACA GGTTCATTTT
ACTCAGTCTC TCTCTACTCA ATGAGGGGCA CCGAGTGTCT CTAGGGCAGG CTGGGTGGTG
15 GTCCCCTAG TATGAGCTTC TCTTACTGTA CTCTCCGGGA ATGTTAGCTT TCTATTTTTC
AGCCTGTACC ACTGTCTTAG GCGAGCTGGC TTCCGCAATG GCGCTGTGG GTCCACAGCA
GGTGTGCTGC CCGGAGGGG CAGCGCTTCT TCTTTGATCC TCTTTCTTTA ACAGTGACTT
20 GCGCTTAGAT CTGGCAGGA AGCTTGCTTT TAGCTTCACC ACCAGAGAGA GAGGTGACA
TGACTTCGC GCGCCCTCAC CAGGCTGGG AACAGGGGG ATGTGGTAG AGCGAGTTTC
CTCTGGGCTC CTCAGGGTG TTTTCTACTA GTCACTACTG TCTTCTCTTT GTAGCTAATC
25 ATCTCATATT CTTCCTTTC CTGTGGGAG TTGGAGATG CTCTGGGTG TAGCTGTGAC
CTGCGACTG AGTTGGGAA AGAGATAT CAGTGGGAC TGTCTGCTC AGAGCTCTTG
30 ATCTACGCA CCGCTTAGA TCGAGACTG GGTCAAGCT GATGAGAGC AGCGCTGGC
AGCAAGCTG GAGTGGCT GAGTGGGAG AGAGCTGAC TTTCTTTTC CTCTCTCTCC
TCCACATATA CTGAGCTCT ATCTGTTAG GATCTTTCTA GCTTTTTC CTCTGGGTG
35 GACAGAGGA CAGAGGAA GGGAGGCTC AGAGAGCA GCGCTTCTTT GTCTCTGGG
GTAAATGAGC TTGACTTGA GTAAATGGAG AGACCAAGG CCGTGTATTT TTATTTTCA
40 TAAATGTTA GAGTATATA TATATATATA TATATTTCTT TAAATTTTG AGTCTTTGAT
ATGTCTAATA ATCTATCC TCTGCTCTCA AGCTGTAGT AGAGATGA AGAAACTGT
45 GTTTATTTA AGATGTTAA TTAATGATT GAAACTGGC TGTGGCTACT GCTTCTTAT
GTGGGGGA CAGGCAATG GTCTGGGCC ACTTTTGA GGGAAATGT TTTGGCTGT
GACACATTT GACCCAGTA TGGGCTCTT CTGCTTAGTA CTGCAAGG GACTGTTAG
50 GTGTCTGTC CACTCTTAC CCGCAGCCC CCHTTAGGG TAAAGGATC CCGAGCTAG
GTGAGGGGCC AGCACTGCC TCACATGAG TTTCTCTCTG AGATGCTCA GCTCAATCC
AGACACTTG TTGAGCTTT TTATTTGAT TTATGAGCT GATGGGATTT TGACTGAGT
55 GCGTTATGGA GAGTATCCC ACCCTCATG AGACAGAT CACTCTCTGC CATCTATCT
GCGTATGCT ACGACAGC AGCTATTTA GGGAGTGGC CAGCTTACT GAGTACAGG
60 AATTTCCAG AGCGCTGGGG CAGCGCTGAG NCCGAGTGC CAGGCTTAG AGTAAAGGA

CATTCGAAA GTTCCAAAC ACTTGGTAG CCACTAGGTT CTCTGTTTTC CCTTCCCTTT 1740
CCTTTTCAAA TCCGACAGTT TCTGTGTGG GAGAGCTGT AATTAGCTTA GTCCAGTAC 1800
CAGATCCGAG CTAGGGGGC AGCTGNTTGG GATAGCTCCA AGAAAACCTG GGCACAGTA 1860
TTTTTCCATT TATNAGGACT GTGCGATAAA TTTTAAATG AGTTATATG AAACAGATT 1920
TCTCCAGCTC CAGGGGAG AGGTAGGGC TGAGTCCCT GCTGTGGGCC AGCGCTTTT 1980
AGGGTGTGCT CTGTACTGC ACCGAGCAG GATGATCTG GTTCTGGGG AGGTAGCT 2040
GCGCTTGGC GAGTTCTGCA CCGATTAAG AGTCCAAAC CCGTCTTTC GTGTCTGAG 2100
AGATGGTTAA ATGGGTGATG GATGAGCAG ACTGAGAGTA CAGAGATGA CTGATGTG 2160
GAGAGGGGG GAGATGCT GGGCTGGCTA GCTAATGTC CCGCTTTTCA GCGATTTACA 2220
GGAATGAG CCGAGCTTGC TATGAGATT GTTTCCTTC CACTGTGGA TCGACTCTC 2280
AGAAATTTG AGTGAAGCT GCACTTCTC GAGACTTTC TTCTTGGCT TGAGCTCTC 2340
ATCTGTTTGG CCGTTTCTAT AGCGTTGAC AAGAGCTGC TCACAGGAG CAGAGCTGC 2400
CGAGAGTTC AGAGTTTCAA CTGGCGTTT ATCGCTTCTA TAGAGCACA CAGAGATG 2460
CCTTGGGACT GAGCTCTCT CACTTCTG GTTTCAGT TCGACAGAC CACTAGAGC 2520
CTGTCTGCA GTTCTCTCT GGGGAGAAC TGTACAGGC CGCTACGAC AGTCCGTGT 2580
TGAGTTCGC CAGGTCTAT CTCTCTTACA TAGAGCTGT CTGATCTG GTCTCTCTC 2640
ACAGTATGA TTTTCCGAC AGCGATATC AGCGGGATG GATGAGCTC CTCTGTCT 2700
GATTTCTTGG CAGGCTTGG GCGATTTGT TCTGCTTTGA GCGATCTGG TAGGCGAGC 2760
TGGCAGTTT TTTCAGGGCA GGGGTATTA ACTTTTCCG GATGTGATC AGCAACTT 2820
TGAGTGGAC TTGACAGAA TTCTTCAAT CTCCAGATG TACAGCTCA CAGAGATGT 2880
CCTTTTCCAG GTCACTTAA GCTGTGTAG TTTTGTTC ACCCGATTC TCATCTGTA 2940
GATTCGAAA CTGGACTTA AGGGAAAA GCGATGCTT GATGAGGAC AGAACCCAT 3000
TGTCTTCAC ATTTCTGGC TCACAGAGG CTTTCTTAG TTTTCTTTC AGATCTCT 3060
TCCGATCAG GAGATCANT TTGGACTCT CTCTGAGTA GCTCATTTG CTGCTGTTA 3120
ATCTGAGAC CATTAGATTC ATAGATGGA CCGTTTTTGA ATAGCGAGT CAGAGAGT 3180
ACTTCTGTC AAGGTGAAA ATCTTCTCT GATCATGCC TCCAAATGG CACTCTACT 3240
TTAAATATC TTGATCCAA GCGTGCATC CCGGTATTA GAGGCGATC AGCAAGGCT 3300
GCTGCACTG CTTHACGAC TGAGCTCAG CTTTCTTGA ATCTGTCTGT GTAGCAGG 3360
AGAGATCT GTACATCT AGTGTGACT CTTTGCTAG CTGGTATTA GTGCTTTTA 3420
TGAGCTTGA CTCTCCAG GCGACGCA TGTCTCTAG GATTCCTTG ATCATTTCT 3480

5 CATTGATGCT GACTGCGAGT TCTAGAGGT CCATGCGGG TTTTATTTTA TCGAGTATG 3540
 CTTGAGGCT CCACACAGA ATTGTACTT CACACCTTC CTTTAGAG TCTGCAATCT 3600
 TTGACATGG CACAACTTA GCGACATGT GTTTGCGGT GTTTGCGGT CCGCAGTAA 3660
 TTTTATTTT CCGCTCTTC AGTATCTCT TCACTTCTT TTTCCCGCA ACTCTCTCA 3720
 10 GTTTCGGGT GATGAGTGC AGTTCTCTT CAGGCTGCG AGGTTCCCG ATGCTTCGGT 3780
 ACCCTGCTT CCGCGCTCA CCGCGGACC AGAGCGCTT CTTGCTGAC CTGCGCGCC 3840
 GCTGCGCGG AACTGTCAG CGAT 3865
 15
 20 (2) INFORMATION FOR SEQ ID NO: 290:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1910 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 25
 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 290:
 30 AGGAGAGGA GAGAGGGG TGTGCGCG CCGCTTACC AGAGCGAGC GAGCGGAGC 60
 AGGAGTGG CTGTCCCGA GCGAGCGCC GAGCGAGCC CCGCGCGCC CCGGAGGAC 120
 GCGCTTCCA GCGAGCGCA CTCCTAGAG GAGGAGAGC GCGAGAGAG CTCAGCTTC 180
 ACCAGCGCC CTGCGCGCG CCGCGACTT CCGAGGCTC TCGGAGCTG CCGGCTCTC 240
 CTGCGAGCT CCGAGCGAC CCGCTGTCA GAGCGCATG AGTTGCTCT GTCTCTCTC 300
 TCGAGCTGC TGGCGCGCT GTCTGCGC AGTCACTG AGAGAGAA GGAATGAGC 360
 CTTTTCATT ATGATTTCA GAGCGTGG AGTGGGCGC TGTGTTTCC TGTGCTCTC 420
 TTCTGCTTG GATCTCTCT TATCTTACT CCGAGGTGA AGTGCATTT CATTAGAG 480
 CCGCGCGCC CAGGAGTGA GAGCGCGAG GTGAGAGCC TATCTACCC CATTGACCA 540
 GAGCGCGCA AGGAGAGAA CTGATGTCA GCGATAGGT GGAAGCTCT GGAAGTGG 600
 GCGCTCTCT GAGCTTTTG ATGCAATGT GATGCTTAA GAAAGCGCC CACTTCTGA 660
 ACAGCGCTT CCGCAGAGA AGCGAGAAC TTGTGTGTC CCGAGCTAT CCGCTCTAC 720
 ACATCTCTT CACTGTGTA TCGATCTAC ACTTCTCTT CACTGAGC CTGCGCTCT 780
 GCGAGCTCT CTGATGTGT GTGTGTGTG GTGTGTGTG GACTGTGTG GTTGTCTAC 840
 TGTGCTCTT GTGCTACTT GTTGTGTGT GATATGTGT TGTATGTA ACTGTGACT 900
 CCGTTTCCA GCGAGGCTT GAGCGAGAT GCGATCTCT CTGCGCGCC CCGCTGCGC 960
 60

TCGATGCT TCTGCTCTA GAGCGCTCT TGTGCGCA GAGCGCGCC CTGCGCTAT 1020
 TTAGGATCC GTAGGTAG AGCAGGCGA GTGTCTTCA GTGTCTTGG GAGCTGCGAA 1080
 5 GTTTGAGC ACTTGTCTT GATCTCTAT GAGCTCTTT CACTCTTTA ACAGAGCT 1140
 TGTCTTTTA TCGAGCTGA TCGAGCTG AGGTCTCTT AGGAGCTGA GATGAGAGC 1200
 AGGAGCTGG TGAAGCGAG GTTGAAGTA GCGAGCTAT CCGCTCTCT GTTAAATTC 1260
 TTTGAGGCG CTTGAGAG GAGTCCCAT CTGCGCGCC GTTCTACAA CCGCGCGCG 1320
 ATTGAGGCC CAGGCTTCT ACTTCTGCC TGGGATGT GTGCGCTGA TATCTCTCA 1380
 15 GCGATGCT CATGCGCTT GAGCGCTAC CCGTCTGAC CTTCTCTCT TCTGAGACTT 1440
 CATTCTAG CCGAGCTAT CCGATGAG ACTAGAGTCT CTGATGTG GTCTCTGCA 1500
 GCGATGCTT GAGAGCTCC TTTTCTGTT GCGCGAGAC ACCGAGTGG ATGAGGAG 1560
 AGGAGGCC TTTCTCTCT TCGTACTCT CCGTATGAT GCGAGAGAG CCACTCTCC 1620
 CATCTTCTG TCTGCTCTG GTGTCTGAG CCGGTGAGC AGTGTGCTT GAGACTGAG 1680
 25 AGGCTCTCT GAGCGCTTG GAGAGTGG AGTGTGAG TCTATAGAG AGTGTGACT 1740
 CAGCGAGT CCGCGCTCT CTGTCTCTG TGTGCGCG GAGAGAGC AAGCTGTGG 1800
 CTGTGAGCA TTTCTTCTT CTGTATGTT ATCTATCTT AAGAGAGCA GAGAGAGCA 1860
 30 ATATATATC CTTTCTTCTT TAAAGAGAA AAGAGAGAA AGGCGCGCG 1910

(2) INFORMATION FOR SEQ ID NO: 291:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3776 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

CCGAGCTCT TTTGATGCT CCGTCTCTT CCGCTGCA CTGCGCTCT CAGCTGCGG 60
 ATGAGAGCC AGCTTGTTC CCGCGAGCC TCGTCTCTG TGGAGAGAG CAGAGCGAA 120
 50 CCGCTCTCT GTTTCTCTG CCGTCTCTG CCGCGCTG GAGAGCGCA CCGCGCGCC 180
 AGTGTCTCT TTTGAGCTG TGTGAGCTT TACCGCGCG CCGCGAGCC CCGCGCGCG 240
 GAGAGCGCC CCGCTCTCT TCTGAGCGT ATCTAGAGG TGTGAGCTA TCGAGAGCA 300
 55 TTTCTAGCA GAGAGAGCT CCGATCTCT TATTTCTTA TGAAGCTAT AATGAGACA 360
 AGCTGTCTT GTCTCTCT CAGCGAGAG TGTGATGAT GAGTGTCTT CAGCTGACA 420
 60 CAGTCTCTT GAGAGAGAG AGAGAGAG AGCTGTCTT CATCTCTCT TCTGATGTA 480

540 CTTCTCTCTGA TCGAGAGAT CCGATGAATA GAGTTCTTTC GAATTAACCTT COTCTAGCC
5 TAGGGAGAT CATTAGCATC CAGCCATGCC CTGATCTGAA GTACGGGAAA CATTATCCATG
600 TCGCTGCCAT TGAATGACCA GTGAGAGGA TTAATCTGTA TCTCTCTGAG GTATACCTTA
660 AGCCCTACTT CATTGAGAG TATCTAGCCA TCCGGAGAG AGACATTTT CTTCTCTCTG
720 GTGGATCTGG TCGCTCTGAG TTCAAGATGG TCGAACACGA TCTTAGCCCT TATTGCAATG
780 TTCTCTCAGA CACATGATC CACTCTGAG CCGACCTTAT CAAGCAGAG GATGAGAG
840 AGTCTCTGAA TCGATGAGG TATGATGACA TTGCTGCTG CAGGAGAG CTAAGCTAGA
900 TAAAGAGAT GTTGAATCT CCGCTGAGC ATCTGCTCT CTTTAAAGCA ATTCTCTGTA
960 AGCTCTTAG AGGATCTCT CTTAGCAG CTCTCTGAC AGGAGAGAC CTGATCTCTC
1020 GAGCTCTAGC AATGAGACT CAGGCTCTCT TCTCTCTGAT CAATGCTCT GAGATCTGA
1080 GCAATCTGCG TGTGATCT CAGACCAAC TTCTTAAGC CTTTGAAGG CCTGAGAGA
1140 ATCTCTCTGC CATTATCTC ATGATGAGC TAGATGCCAT CCGTCTCATA AGAGAGAAA
1200 CTCACTGCGA GTTGAAGCG CCAATCTAT CAGCTCTAT GAGCTCTAT GATGCTTA
1260 AGCAGAGGCG AATGATGAT GTTATGAGC CAGCAGAG ACCGAGAGC ATTTAGCCAG
1320 CTCTACGCGG ATTTGCTCG TTTCAGGCG AGGTAGATAT TCGAATCTCT GATCTAGAG
1380 GAGCTCTAGA GATCTCTAG ATCTATGCA AGACATGAA CTTGCGAGAT GATGCTGACC
1440 TCGAACATTA CCGATGAGA CTGACGCGCA TGTGCTGCT GACTTAGAG CCGTCTCTC
1500 AGAGGCTCT CTGAGAGCA TTGCGAGAA GATGATCTC ATTTAGCTAG AGATGAGAC
1560 CATTGATGCC GAGCTCTAGA ACTCTTAGC ATTACTATG GATGACTTCC GTTGGGCTT
1620 GAGCGAGAT TACCGATCG CACTGCGGA AACCTGCTA GAGCTGAC AGGTAACTG
1680 CGAAGACATC GCGGCGCTAG AGATGCTCAA ACTGAGCTA CAGGAGCTG TCTAGTATCC
1740 TGTGAGGAC CAGACATAT TCTGAGTT TGGATGACA CTTTCTAGG GATTTCTCTT
1800 CTATGAGACT CTTGCTCTG GAGAACTTT GTTGGCATA GCGATCTCTA ATGATCTCA
1860 GCGCACTTC ATCTCTCTA AGGTCTCTA CTTCTCTAC ATTTGCTTTT GCGATCTCTA
1920 GCGCACTTC AGAAGATCT TTGAGAGCG CCGCAGCT GCGGCTCTG TCTATTTCT
1980 TATGAGCTG GATTTGATG CAGAGCTG TCGAGGTATC ATTTGAGAT GTTGTGGGCG
2040 TCTGAGCGA GTCTGACG AGATCTGAC AGAATGAGT GCGATCTCTA CAAAAA
2100 TGTGCTATC ATTTGCTCTA CAGAGGCTG TACATCTAT GATCTCTCTA TCTCTAGAC
2160 TGGGCTCTT GATGAGCTA TCTATCTCC ACTTCTCTAT GAGAGTCTC GTTGTGCTAT
2220 CTTGAGGCT TACTGAGCA AGTCTCTCT TCGCAGAT GTTGAATCTG AGTTCTCTG
2280

2340 TAAATGACT AATGCTCTCT CTTGAGCTGA CTTGAGAGC ATTTGCGAG GTCTCTCTGA
2400 CCGTGGCTC CTTGATCTCA TCGAGATGA GATTAGGCGA GACGAGAGA CCGACAGAAA
2460 CCGATAGCG ATGAGGTAG AGAGATGA TCCAGTCTT GAGATCTCT GAGATCTCT
2520 TCGAGAGC ATGCTCTTT CCGCTCTTC TGTCTATGAC ATGATCTCT CCGATCTCTA
2580 GATCTCTG CAGACCTTC AGCAGATCT GCGCTCTGCG AGCTCTGAGT TCGCTCTG
2640 GAGCAGCT CAGCTCTG CCGTCTAGG CAGTCTAGG CCGCAGAGT CCGATCTCTA
2700 CAGCAGAGC ATGATGATG ACCTCTATG CTAAGTCTG GTTGGCAGG TCGATCTGAG
2760 TCGCTCTCT GAGCTCTCT CCGTCTGCT GCGGCGCTT CCGCAGAGC CCGACAGG
2820 GTGCGCTAC AGCTCTCTC ATTTCTCAT CTGAGAGTT CAGCTAGCT CTGATCTCTG
2880 ACAGGCGCT TCTCTCTCA AATAGAAA CAAAGCGAT AATATTAAG CCAATTTCT
2940 TTGTAAGCG CAGATGAT TACAGAGG GAATGCGCC TTGCGCTATG CCAATTTCT
3000 TGTATTTG CCGATCTGAG CCGCTCTG TCGCTCTG ACGAGGAG TACTCTGAG
3060 TCGCAGTA AGCATCTC ACTTCTCTA ATCTCTGCG AGCGCTCTT TCGCTCTATC
3120 CAGCTCTCT AGCTCTGAG CCGCAGCT TCTCTGAT TTAATAGAG AGTCTCTG
3180 TTTATTTAC ATCTCTCTA GTTATCTG GAAACTAT CAGAGCTT TCTTAAGCA
3240 AAAAAAGCA TTTCTTAAA GGCATAAA CTTTGTCT AATGCGTA AAAAAA
3276 AAAAAAGCG GCGCTCTCTA AAGNCCAN CTCTCT

(2) INFORMATION FOR SEQ ID NO: 292:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1695 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(41) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

50 TTGCAATGTT TGAATCTCC TCTCAGGCC AGCTTAGAG AAGATCTG TACTCTCAGA
120 GTTGAAGAG CAGCGCGAG TTCTCTGCG GTAGAGCGCG AGCTGAAGT ACAGCGAGG
180 CCGAGACAC GTTCTGTTGG CCGAGAGAG GCGCTGCGCC CAGGAGAGG AGAAGACCT
240 TCGAGAGAA CAGCAGAG CTGAGTCTT GTGAGAGAG GGAACAGT CCGCGCGCG
300 AAGAGGCTT CTGGTGAAG ACCCACTCG GCGTCTCGCC CTTCTCTCTG CTGACATCTG
360 CTTTGGCGG AGCTTGGGG ACCGCTTGG CTAAGCATT TGAATCTCTT TTTGGTGA

5 CCGCGCTCTG CCGCGCGGCC TGTGAGTTGA CTTACGCCCTT GCACACCTAC CTTAGCGAG 420
AGAGGTTGTA GCGATGCTCAG AGAGTTGCA GCGCTTTTTC ATTGTGTCAG TTGTGCGATG 480
10 ATGGAMTTGA CTTAATGGA ACTAATATG AGTGGATC TCAATGTACA GAGCAATTT 540
CCCAATCTGA TCGACATAT GCTTCCATC TTGGTTGCA GATTCAGCTG CCAATCGCTG 600
AAGTGAGCA AGAGCAACTT ATTCGCTGA TCGCAAAAT GCACCTACTC TTTCCTCTTA 660
CTCTGTGCG AGCAATCTG AGTGCATCA TCGACTCGC AGAGCTTC ATAACTCTT 720
CATGCACTTT TTATCTGCA GCGATGAGG GAATAATAGT TATATTCGAG TCTAGCGAG 780
15 AAATCGATTA GCGACCAAT TTGAGCGAG AGCTTACAAA TTTCAGAGAA TCACTCTTA 840
GGAATATGTC CTATCTGCA ATGAGAAAT CAGAGGCA CAGGAATTTT CTTGAGATG 900
GAGAAATGCA TGGCTTTTTA AGATCCCTCT CTCTTAATC TGGTGGAT TTAACTACAA 960
20 CTCTGTCTCT CTGGTGATG GTATTCCTTT GCAATTTGTTG TCGAACTGTT GCTACAGCTG 1020
TGGAGCATTA TTTTCCCTCT GAGAGCTCA GTATCTATG TGACTTGGAG TTATGCAATG 1080
25 AACAAAGCT AAACAGATAT CAGCTTCTTT CTCTTGTGT TGTAGATCT AAACAGAG 1140
ATCATGAGA AGCGGCGCT CTAACTACAA AAGTGAATCT TGTCAATCT GAAATTTAG 1200
30 CATTCTCTTT TTAAAGACA AGTGAATAG AACTTAAAA TTGCACTCTT CAGAGAGCTT 1260
TTAAATGCTT TTCAATGAT ATAGGCTTA AGAATCACT ATAAATGCA AAATAGTTA 1320
CTCAAAATG TGAAGACTGT ATTGCTATA ACTTTATG AGTGTCTTTT GTATTAATTT 1380
35 AAGAGTGA TTTTGGAT TGTATTATA TTTTACTAAT ATCTGTAGCT ATTTTGTTTT 1440
TTGCTTTGCT TATTGTTTT TTGCTTTTC TTAGCTATGA GCTGATCATT GCTGCTCTTC 1500
40 AGCTCTGCG ATGATACTGT CAGTTACTT AGTTACAG AGTAATATTT AGTAGAATG 1560
ATGCTTTGCG TCGAGATG GCGAATATC TGTAAATTTA AATTAGCAG GAAATGACTT 1620
TTATGCAAC TACTTTTCA GAACTGAA TCAATTAAT TTATTTGAA TATTTAATA 1680
45 AAAAAAAAAA AANCT 1695

50 (2) INFORMATION FOR SEQ ID NO: 293:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1501 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

60 CACTTTGAC AGCTCTTGC TCTTTTGT TCACTTCAA ATAGCCCGAG GAGTGGGCTT 60

5 TATCTCTCAA TATCGACAT CTAAGCTTC TCTTGCGGGA TGGGATTTG GATGCGCAGA 120
ATCTGTTTGG GATCTCGCGG TTATTTTCAG TGGGTGAAA AGCAGAGCTG GCGCTTTGCG 180
20 TCTCTTATCC CTGAGGCTGG GTAAAGAGA CTTATCTAC ACCTGTCTTT CCGTACTCTC 240
TCTTTTGTTA GCGAGGCTTC ATTCATAGTT CCTCAAGAGA GTCTTGGCT TAAAGCTGTA 300
10 GCAAGGCTGT GCTAGCTGGG GATTTTGGAG CAATAAGCTC GATTAAGCAT GATTAAGCTA 360
TGGAGTGGCG CTCGAAATTC AGACAGAAAT GCGTTGAGAA GCGCGAGGGG AGCATGCTCG 420
TCTCTGAGTG ATAGATGATG GAGAGAGCT CCGTACCTTG GAATAATGGA ATTGAAGGGG 480
15 TTATGACAAA ATAGAGTGGC TATGTGAGGA TTTTCCAAA GTTTTGTCCA ATCTTATCAT 540
TATTAATTTT TATTAAGCAC AGAGACAAAC CAGAAACGGA ATAAATGTTAC TTTCGATGCT 600
20 TTATTTTTTT GTTCTAGCTG TGGCTTTGTA CATTGAGAG AATGCTATAT CTTGCGCATTT 660
TTGCTTTTAA AGCTTTACCA CTTTCCCAT TTATGCTTAA TGGGAAGATA CAGATGTGCA 720
AGTCTGCTTT TTGTTTTTTT GTTATTAATTT TTTTTTTTTT GCTCTGTGTT ATGACATATT 780
25 TCGAGATGCG AGAGATGCG AGAGATGCT CTTTGGAGCC NATGTGTGCA TCACTAGCT 840
GCACTACTTA TCACTATGCG TCACTGCTGT TTCACTCTTA TCTCTCTCTT TTCACTCTTA 900
30 TTTTATTTTG AATATCGAG AACTATGCG AATGCAAGCG TCACTACTTT GCGAGATGCG 960
TATCTCTCTT TGAATGTGAT AGTCAATGGG TCACTATGCA TATCAATATC AGCTCTGCTT 1020
35 TTTGCTTTTA ATGTTAACTA ATGAATTTCC AGAGATGGCG CTTAGAATAT TATTTTAAAG 1080
ATTAAACAGG AGCTTCAGAA AGAATGAGA GCGATGCTTC CTTTNGCCTT GCACTACAAA 1140
AAGAGAGAG AGACTGTCTT GTTGTAAAGC TTTTCAAAA ATCTGATATAT GGTAAAGTTAC 1200
40 TTGAGAGCCT TCAAGAGAT GTCAATCTTT TTTTCTGTGT AACTGAGAA CTTGTGTGAC 1260
CATTAAGCAT GTTATCAGCT TGTACTGCTC TCAATAGCTT GGTATTGGA GAAATATTTG 1320
GAAATTTTGG CTGTGTTCTG TGAATATTAAC CTCGCCAAA TATTAATGTA CTGTTGTTTC 1380
45 TACTTGTGTA TTTCAGAGCC TTTTAATTAAC GCAATTAATTT CTGTGTTCTT AAACAGATTA 1440
AATAGTTGTA AGTTTGGCTG CATGATGGA AATTAATTAAC CTGTATCTCT GTTAAAGAAA 1500
50 A 1501

55 (2) INFORMATION FOR SEQ ID NO: 294:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

5 TGATGTGGT CCGGGGTGCG GATTGGCGCG GCGTCCGCGG CCGCTGTGCG TTGTCCGCGC 60
ATGGCACTGT CCGCGGGGGT CCGCCCGGCG CTGGCTGAGG CGGTGGCGCG GCGCCGGGTR 120
CTGGTGGTGG GCGCGGGCGG CATGGCGTGC GAGCTCTCTCA AGAATCTGCT GCTCAACGGT 180
TTCTCCGACA TCGACCTGAT TGATCTGGAT ACTATTGATG TACGCAACCT CAGACGACAG 240
TTTTTGTTC AAGAGAAACA TGTGGGAGA TCGAAGGCAC AGGTTCGCGA GGAAGTGTGA 300
CTCGATTTT ACCGAAAGCG TAAATCTTT GCTTACGATG ACAGCATCAT GAACTCTGAC 360
TATATATGCG AATTTTTCG ACGATTATTA CTGGTTATGA ATGCTTTAGA TACAGAGCT 420
GCGCGAACCG ATTATTATAG AATGTGCTG GCGAGCTGATG TTCTCTTTAT TGAAGTGGCA 480
ACAGCTGGGT ATCTTGGACA AGTAATCTACT ATCGAAGAGG GTGTGACCGA GTGTATGAG 540
TGTATCTCTA AGCGGACCGA GAGAACCTTT CTTGGCTGTGA CAATTCGTAA CACACTTTCA 600
GAACTATAC ATTGCATCTG TTGGCGAAG TACTTGTTC ACGAGTTGTT TGGGGAGAA 660
GATCTGATC AAGAGTATC TCGTGMAGA GCTGACCTG AGCTGCGCTG GGAACGACG 720
GAGCGGAG CGAGGCTAG AGCATCTAT GAGATGCTG ACATTAAAGG TATTTCTACT 780
AAGGAATGGG CTAAATCAG TGGATATGAT CCAATTNAAA CTTTITAGCA AGCTTTTAA 840
AGATGCAATC AGTATCTGT TGACATGGA CAATCATGG CGGAAGAGA AACCTCCAT 900
TGGTGTGAC TGGGCTGAG TGAAGTCA AGGAGAGAA ACGATGCTAT CAGATCACA 960
GAATGACCC CAGTTAGGCC TGAAGACCA GCAAGTTCTA GATGTAAAGA GCTATGCGG 1020
TCTTTTTC AAGAGCATG AGACTTTGAG AGTTTATTTA CGAGAAAGG GCGATGAGC 1080
TGAGCTGATA TGGGATAGG ATGACCCATC TCGAATGGAT TTGTCTCACT CTGCTGCGAA 1140
CCTCAGGATG CATTTTTCA GTATGATAT GAGAGTAGA TTGTGATATCA ATTCATGCGC 1200
AGGGAACAT TTCTCTGCTA TTGCTACTAC TAAATGCGTA ATTCTGGGT TGATATGAT 1260
GGAAGGATG AGATTTTAT CAGGAAAT AGACAGTGC AGAATCAT TTTTGATTTA 1320
ACAGCAAC CAGAAAGA ABCTTCTGT GCTTGTGCA CTGGATCTC CGAAGCCGA 1380
TTTATATTA TGTGCGACA AGCGAGGT GACTGTGGG CTGAATGTCC ATTAAGTGAC 1440
TGTTCGACC TTACAGACA AGATATGAA AGAAAAATTT GCTATGCTAG CACGATGT 1500
CGAATTTGA GATGGGAG GAAATCTT AATATCTCC GAAGGGAG AGAGCGAGC 1560
TAAATATC AAGATTTGT CAGATTGCG AATTAGAAAT GCGAGCGCGC TTCAAGCGA 1620
TGACTTCTC CAGGACTATA CTTTATGAT CAAATCTTT CATATGAG AGCTAGGAA 1680

GCACCTTGAA TTTGAGTTG TTGGTATGC CCGGAAAAA GTGGGGSCA AACAGCTGA 1740
AGATCTGCC AAGACATA CCATGSCCA GTATGATGG AGCTGAGCC TCCACTCCA 1800
5 CAGCTCAGA GCAATATAC GTTCTCATAG TTGATTCGGA TGAAGAGAT TCTTCMAATA 1860
ATCGGAGCT CATGAGAG AGAGAGCG CAGAGGAAA TTAGATGAGA AAGGATCT 1920
CAGTCMAAG AGTCACTTA TAGACAGAA GGAAGAGCTT GATGATGTCA TAGCATAGA 1980
TTGACAGGAA ATGCTCTTAA ACGAGCCT CTACTATTT AGTTATCTG GCGAGACCA 2040
GATTTTATG TCTTTTGTTC CAAAGGAAA AATGTGACAG CAGTCACTTG AAATGATTC 2100
15 TCGTCCCTTT GAAGCATTC ATTTTCTAG AACTGTAGA CAGTTGCGAG TATGCTGTAT 2160
TGAAGTAGG AATATATTT TAAAACTCT TTGACMAAG TGTGTGCTATA ACCATGCTAG 2220
AGATAAACA ACACATGCA TGTGCTCTT TTATGTAAA TACCTTAGG TATCATTAAT 2280
20 AGTTTCAAA TATTTGCTT TAGTAAAGT GATACCTGCT TATTAATAT ATGCTTTAT 2340
TTTTGCTAG AGAGGATTT ATTTTATGCC TAGATCTAAC CATTITGATA CTCTTACTG 2400
25 ATTGAACAG ATTCAAGAA GTATGAGTG CTATGCTATG AACTGTGTT TTAATGTGA 2460
CATGCACTA TGTATTTAA TGTAAACAA TGTTAATTTA CTCAGTTTT CAGTTGTAC 2520
CGCTGTAT GTCTGTAA GAAGCAAT TTGTGTAT GTTACAGTTT CAGTTATTT 2580
30 ATATGATG TTTGTAAA CTCAAATAC GACTATACT ATGACCAAA TAAATGCGAT 2640
TGTCTCTG TTAATAAAN NACGAAAA AAAAAAACA AGA 2683

(2) INFORMATION FOR SEQ ID NO: 295:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1454 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

50 GCACTGGGG TGGCTCTTAG GCGCAGGAT AGGCTGGGG AGCGCGGCC TGTGGCTG 60
ACAGAGCTT TCTGTGCGAG GTTCAAGCC CATGAGGTG GTCACTGCT TGAAGCGGA 120
CAGCTACTG AGCACTGCT TCTGACGA CTTGATGCTC GTGCTGCTT CTCTGGAAG 180
CAGCGCTCG CCGGAGCTG TTGACAGGA CTTCTACTC GAGTTTGGG AAGAGACAC 240
55 AGGGAAGATG GAGACTAG AGCTGATCA CTTTATGCG GTCAATTTTA CTTACCGCAG 300
TGAAGGAG ATTTGGAGC TGAGTTTAC TGTGGCCAA AAGATGGCTG AGCGAGAA 360
60 GCGCCGACC CTCAGATCC TGTGTACT GCAAGCTTC CAGTGGGA TCGACGCC 420

5 TGGTGTCTGC AGGGGCCCC TGGCCGCCAA GACACTCTCG CTGACAGCT CCGAGATCTT 480
CCTCTGTGAT GAGGAGCTGT TCGACTACCC ACTGCCCCGAG TTTCGCCAAG AGCCGCCGCA 540
GAGAGACAGG TACCGGCTCG AGAGTGGCCG CCGGCTTCGG GACCTGGACC GATGTCTCAT 600
GGGTACAG AGCTACCG AGCCCTCCAC CTGGTATTCG ATGAGCTGCA AGGTCACTAC 660
10 CTGATGGGCA GTGTGACCTT GAGCAGCTT GGGGAGGTGC CAGGTGGCCC GCTTAGAGCC 720
AGCCAGGCC GTGAGTCA GTGGAGTGT TTGTGCCCA GTCTGAGAG CAGAGAGAG 780
CTCATCTGC TTGTGGCTCG CAGTGGAG GCGCTGTGTG GCTTGGAGCT GCTGTGAGCC 840
15 TACCGGCTA GCGCAGGCA CAGCCAGCTT GTGTGTGCA GCTTGAGCCC TACTGGGGCA 900
GGGAGACAG GTTTTGTCTT CAGTAAAT GTTTATCTCT CCGTTTGGTA CCGTAAATTTG 960
20 ACTGTCTCG CAGAAATGT AGATGTGTG TGTGTGTCTT TAAATCTTTC TCAATTTGGG 1020
AGTGAATAT CCGGGCCCTT CAGGCTGTCT CCGTGTGTG TCAAGCTGCC AGAGTGTGTA 1080
CAGCTGTCA CACAGTGTG GTGTCTCTCG TTGTGGGACC GTTTTAAACA GTTGAGCTG 1140
25 TGGTGTGAC TTATTTCTCT AGAGTCTTT TCGTGAATGT TCGATGTGAG TCGTTTGTG 1200
CTGTGTCTGT TCGTGTCTGT GTTGTCTGTG GCATCTTCTT GCTAAATCTG AGCTGTGTAG 1260
30 CAGATGTGAC ATTTGAGCTT CCGACCCCAT ATTGTCTTTC AAGTGTGAG TGTCTCTGTA 1320
TCCAGACAG TCGGAGACC CTTGGGGCA GGGAGCTCG AGCTGCCAGC ACCAGCTGTG 1380
ATCTGTCTG CCGTATTTCT GTATTCAT ATAGGAGAT TTGAGAGCTG TAAATATATA 1440
1454

(2) INFORMATION FOR SEQ ID NO: 296:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

50 ACCCTGGCAT GCGCCAGAA CAGATACCA CCGAGCTTAC ACAGCATTA ACTCTCTCA 60
ATGAGAGCA ATCATTCAC ATGAGGAG ACATCTAT GATCTTTAA GGAATGTTT 120
CGCTTATGT TTGAGATTA TATCTGGTA ACTCTTAAT CCGATTAAT AGCTTAGGC 180
TGGAGACAA TGGGTGCA TTAGGAGAT TAAAGATCA GTCCAGTAA ATGATTCAT 240
60 AGACTCAT AGACCACT ACATTAACA CTATTTAGG AGCTGAGG CCGCAAGTA 300

5 GAAATGTGC ATATATCTG CTCTGTGTA GCTCAGAGA CAATTCGAC ACAGACTA 360
CATTTACCC TGAAGTCAG CTCAGATTA TACGAGAC ACTCAGAAA ATACTTATG 420
AGGGGGGAG GCTGAGCTG GCGTTGAG GATGTATGA ATTGTGATG AGATGAGCA 480
AGACAGGGG TCTCAGATG AGAGAGCT GAAATATGAG CAGGGGCTG GATCAGTGG 540
GTGTATTAG AGCACTTTC CAGATGACC ATGATGCTC ACAGTCTCT GCGTATGTG 600
GGCAGATGT CCGAGCCAG TGTGTGCCC CAGCCCATGT CAGTTTACAT GTCTCTCAT 660
720 GCGACCTCA AAGGTACTT CTCTGTAAA GCTTTGCCCT GATATCAGCA ATCAAAATTA
15 ATCAGGATC TTTCAGACT GCTGTTTTTT CCTCTTGGT CTTCTATCA CTAAAGCTA 780
TCTATTCAG CTTAGAGCA TAACTAATTA TTTTGTTC TCACTACA 828

(2) INFORMATION FOR SEQ ID NO: 297:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

30 TCAATTTCA TTACTGAGA TGAACATG TGAATGACA TTGTTCAGG TCTCAGTTT 60
ACCAAACTT ACTATACCA TCAATCTTCA ACAGCAGAG TCTGAATTA GCGAACATTT 120
TTTTCCTTT GAGAAAGAG TGAAGTGGG CAGACCTTTT AGTCTGAGG GAGCTAGTGG 180
AAATCTAGC ATAGAAATC ATCGATAGCA GCTTTTCTC AATGTGTGA CTCTCAGGG 240
300 GCTAACTGC TTATAGCTTA GATTTATCT TTACTAGAGA TCTAGCGAT AGTGGGTTA
ATCACTACA TCTGTAACT AGTTATATAG CTTCAGACA TGAAGGAGC ATCAAAAGG 360
420 GATGGAGCA ACCCGAGGA TATCGAAGA GGGATGATG AACCCCTTC CTCTGCGAG
GAGACAGG CCAACAGG CACAGACTG AAGCACTTA GATGTTTAG GAGGAGAAG 480
540 GGGAGCTTT GACAGTCTT TCCCTTTTC CAGTTTCAG CAGTTCTCG CTCTTGCAG
50 CCGTAGGCC AGTAACTTT GCGAATTGC AGATTTTCCC CAGATACTA GAGGAGAGG 600
GACTTTGGG GGTGGGAGG GGTGTGTGT GTTTAAAG CATAGTTAC CTGTTCGAC 660
720 TTTTAAAG TAGGAAAAA AATATGTGG CAGGTGAGC ATCAGACTA AATTTGTG
780 TTTTAAATTT GTATGCTCT TGAATATGT TGAATATG TAGTATAGC AGTGAATTT
840 GATTCCTGT TGCATTAAC ACTATATTT TTTCGAAATG TTACTGTCA AAGGCTCTT
60 CCGTCCCTT CCGTTTCTA TGTACTTCT TCACTATTC TTACTATTC AGCGAGGCA 900

5 TAGGCAATCA AGAGCTAGAG CATGAAACAG GGCCTTTTCC AAGTAGGCTC TGGGTGTCTT 960
AAGCCAGCTT GTGCCCTCTG GTTTAGTAGAG TGTATATAGAG TCCCTGGCAC CTTTCTTTTC 1020
AAATGAGCTT AACAGACAG ACTGACGAA GTTATCAGAT TCTTCATCA GATCAGCTAG 1080
GAGTAGAGAG CCGAGGAGTG GAGGGGCTTC CTGAGTATTT GCATTTGGCT GTATAGCTG 1140
10 AGTTCTTTTC CATTTTACCG AATCTTAGC CAGTTACAT TTACTACGAA AAGCGTAGA 1200
TCAATTCAGC CATGTAGTG CTGGTTGGCA GCGATTGGTA ACGAGAGAA CTGCTCATCA 1260
GCCAAATCT AAGCGTTGCC TTTTAGGAGG CACACAGCAG AAGGACTTGG TCTCTCTTTT 1320
CTGTACTTTC TGTACATGCC GTTAGCTCTA GCATCTCAT CACTCTGGC AGCAAAAGG 1380
GAGCAGTCAAT TCTCTTTTCT CTGCCACCC CCTGCCCTTT GTTACCAACA CCAATTTTCC 1440
20 AGGGGTACA TGAGTTCTG AATTTTAA AATTTTTC GTTTGGTTT TCTTGGGAC 1500
TGATAGTGC TTTTAGCAT GTCAACCC CTTCAAGCT CCGAGCTAG TCAATTTTCT 1560
GTATTTTCT GTTACAGTA TTTTGTGTG TGTGTGTTC TGTGTGTTC TTTGCTGTA 1620
TTATATATG ATGATGAT TGATCTCTT TTTTCCCTAA GCGATATGAA TTTTGTTC 1680
TGTGTATAT TGTGTGTG AATAGCTGA GCAAACTGG GCTGACAGC GTTAGGTAG 1740
30 GCTTGCAAR CCGAGAGAG GCGGTGGAG TGTACTTTC CTTGACAGC TGACTTACT 1800
GAGTCTCTGA GCTTTTCTT CAAATCTTT CAGAGCTCA AATGCCACA GAGCTCTCC 1860
TCTTCTCCC ACTGCCANT GCGGAGCCG GACATCTCT ACATCCACA TGTGTTCCT 1920
CCAGCCCTTC CCAATCCAT GCGAAACAG GTACTTTTG GCACTGGGG CATTCATG 1980
2040 GATGCTTGT TAACTGACA CTTAGCCTTC TCTCTCCCT CCGTCTCTC CCGCAATCA
40 CTTCCTAGGA CAGCCGACT GCTTCCGAG GGTCTCTGTT CCTGCTAAC TCCAGAGAG 2100
CATCCGAGG CTTTGTACA GTCTCTAAT CCGTTCCCT CTTCTTAAGA ATCATATGT 2160
ATAGTAGCTT TCGACATA CAGTATTCAT TGGGTACTC CTATATAT CAGTAGCTG 2220
45 GAATTTGAA GGTGGAGTA GTTAGATCTT TAGCTTTTAT TCTTTATTT TTTGTATTAC 2280
TCTCCATG TGATATATAT TGATCTGTT GCTGGCTTTT ATAACTCTA ACGGAGAG 2340
50 GAGCACTGCC TGAGCTTTG CAGATGTTAA TGAGCACTG TTTTAAATA AAGAGRGAA 2400
NKKCAAAA AAAAA 2416

(2) INFORMATION FOR SEQ ID NO: 298:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 545 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

5 GAAITTCGCA CAGCCATAC TTGGCTCTC CTTGATCTT AAGTCACTT TGTGGCTGT 60
TTCTGACTCA GCAGTACTT GCATTTGCC CAAAGATCA CTAATTCCT CTTAGAGAG 120
CAAAATATG GAATAGTATC TGTCAATGCC TCTCTCATG GTTACACAC TCTGCAACG 180
TGTATATCA GTAAACCA GAGAGAGC TGTGGAAGT GACTCTGGG AACTGCTGG 240
15 ATGTGTTTGT GCAGAAATGT AATAGACATA CAGGTGTTG CATTGATCTG GGCCTCTCT 300
ATGTGATAG AAGGTAAA GGCACATC TCTTTGACT TTGGGAAGT CATTCACAAA 360
GAAGATTTT CCAAGATCT TCTGAAGATT GGTAAAAAT AGCGGTTC CAGCCCTGT 420
20 AATGATCA TTTAGATG CTTCTTACC AGGACGAG AACTGATTA CAGAGTAC 480
ATGAAACAT TCAATCCAG AATTTGCAAT ACCTCAATT NAAATTTCTAC CTATTAATA 540
25 NAAAA 545

(2) INFORMATION FOR SEQ ID NO: 299:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1530 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

40 GGTCTCTCTG GGCATCATAC TTGTCACTGG GTAAACAGTT TCCCACTTA CCGCAGTCA 60
AGCTGCTTGC CAGGCTCTC CCGCTCTGTG AATTTTGGAG GCAAGCATCT TCCAGAGAG 120
TGTGTGCTGG CCAAGGATGT GTGGGGCCCC GCGAGAGGTS CTGCGCTGCC GTCCAGGTGG 180
45 TTGGGCCCCG GCGTGAATCT CCACTCTGTC GAGCTCTCAT TACTGAGAG ATCATGCGGC 240
CAGATGATGC CAAGCTGCC GGCATATGTC ACGGGGGGAC CATCTCTGAG ATGATCGAG 300
50 AGCGAGGCC CATCATCAG ACCCGGATTT GCAACAGCCA GAAACGGGAG GCGTGTGTGG 360
CGCGCTTGGC TGTGTGAG GGCACGACT TCTGTCTCTC CATGTGCATC GGTGAGGTGG 420
GCGATGTGAG CCGGAGATC ACTTACACT CTAAGGACTC TGTGAGGTS CAGGTCAAG 480
55 TGAATTCGCA AAGATCTCTC AAGGTGCCA AAGAGCTGAC CAAATAGGCC AGCTGTGT 540
ATGTGCGCTT GTTCTGTGAG AATGTGACA AGGTCTCTCA GGTGCTCTCT GTTGTGTAT 600
60 CCGGCAACA GCAAGAGAG GAGGGCCGA AGCGGTATCA AGCCGAGAG CTGGAGGCCA 660

720 TCGAGACCA GTGAGAGAC GGGGACATG TCGAGCGAGT CCTCAACCA GAGCGCCACA 540
5 CTGTACGCTA GAGCGATGC AGCTTGATGC AGCTTGATGC GCTTCAGAC TCGACCTGCG 600
AGCGCTTTGT GCGAGGAGAT GTGACATGGA AGCTCATGGA TCGAGTTCGC GGGATCTGCG 660
CTGCACCCA CTGCAAGACC AGCATCTGCA CAGCTTCGCT GAGCGCATTT AATTTTCATG 720
10 ACGAGATCAG AAAAGGCTGC GTGATCACA TCTGCGGACG CATGACCTTC ACGAGCATTA 780
AGTCCATGGA GATCGAGGTG TTGCTGAGCG CCGACCTGCT TGTGCGACG TCTCAGAGCG 840
15 GCTAACCGGC GCGCAGTGC TTCTTCAGCT AGCTGTGCTT GAGCGAGGGA GCGCAGTGC 900
TGGCTGTGCG CAGCTGTGTO CCGGAGACCG AGGACGAGAA GAGCGCTTT GAGGAGGCGA 960
AAGGCGCGTA CTGCGAGATG AAGCGGAGCG GAGCGGCGCA GCGGAGGCT CAGCGCTTGA 1020
1200 CTGCTGCTGC CTGCGACTGCG TGGCTGAGT AGCATGCGA AGCGGCGCGAG TGTCCAGTCA 1260
20 CTTAGAGATT GCGCGCTGCG CCGAAAGCTC AATTCAGATT GAGAGCTGCT GTTCTCTGAA 1320
GTTTTCGATT CAGCATGTTA ACTGTACTC TCTCTGCGAA ACTTACAGAC CAAAGCTTTA 1380
25 TTATATATAT TCGATATCA ATGCTACAGA GTTGTTCGCG GAGCGCGCGG AGGCTGTGCG 1440
CAGAAAGCTT CCGAAATGCT TCGGAGCGCG CTGTAGCGTA TCGGAGGAGC CCGACAGCAC 1500
30 TATATAGCT GATCTGCG TCGGAGGAA 1530

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 997 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

(2) INFORMATION FOR SEQ ID NO: 300:

45 AGGTATGTAG AGACATNTA CACTTACCA ACAAGAGAA GGTCTCTGCG CCTTATATT 60
TAACTATGTT TACAGGGAAT GGTATATG TGGCTTTCGG AGATTTGCTT CCAACATGTT 120
TTGAGAGCC CTAAATCCA ACGTCTGATT GATGCTGTA AGACACTTA TAGCCGAGAT 180
50 GAAACCTCT GGGCAGCTT TCGAGTCCA GGTGTGATCG CTGGCTCTGT TCCCAACGAC 240
CCGAGTACG ACGTCTCAG ACGTACTTC TATGTCCAGG CTGGTCAAGT GCGAGGCTCA 300
55 TGAAGGAGAC ATCGATAGG GTGCTCTTGA TGTCTCTGCG TCTGGAATCC ACCAGCGGCG 360
TATCTGCTTT TATGGGCTG GCGACTGAA TTGATGCTT CAAAGCCATC ACCTGTTGCG 420
60 CAAACATTT GACCGAAGG TAGATGATA TGTCTCTCAG TCGTTAGAG AATACCTAGG 480

5 TTATAGGCGC ATCTATGCGA CTGAACCTTG AGACACACTA TCGACAGCTT GCTACCTGTC 540
GGCGAGAGCG ATGTACAAAC ATGCTCAGAA CTGTCTGCGA CAGTGTGCGT GCGAGAGCAG 600
GGCTTTGCGA TTCTGTGCAAT CCTTTAGGAT AAGAGGCGTG HTATATAGAT GTGGGTAAAT 660
AGATCTTTTG CCTTCGAAT TCGTCTGCGG GTGATGCTG CTGTCTCTCT CACCGCTTAC 720
10 CCTTAGTATT CCTCCACTAA CTTTCTCACT AAGTGAAT GAGAACTGCT GTATAGGCGA 780
GAGTGAAGA GGTATATG TGAGGCACT TCAATTCAGT TGAATGCTG TGAATGCTG CTGGTATGTT 840
TTCATATCTG TCGAGCTGCG GTTCTCTATA ATTCAGGTT TGGTATGCTG GAGGAGAACT 900
15 TTGATGAGAA GAGAACTTCT CCTCTCTTAC TGTTACTTAA AATATATATA GCTCTGATTT 960
CAAGTATAGG AAAAAAAA AAGAAAAA AACTCGA 997

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 301:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

30 TTGAGCGCGA CCGTAGGCGC CCGAGAGAA ACTGCGAGCG GAGGTGACC GCGGAGCGAG 60
35 CATTTCAAT CTGTCTGCTTA GAGCTGCTGC ACCACACCA TTTTGGCTGC AAGGCTGTGT 120
TGTCTCGGA CACTACTTTC TAGGCTTTTC CACCGACTT TCGCAGGCG CTGCGCTGTT 180
GTGAGAAAT CCAATCGAA GATCAATGCG CTGTTACAC CTAGCAGGGA ATATGCGACC 240
40 AAAAAAGAA TTGGGATCGG GCGTGGGAGA ACTGGCGGAG AACTCAAGCA GCGAGATG 300
GAGCATGGA TCGAAAAAT ATTTAAAAAT GATCATGCG GAGCATGCTT TGTTCCTGGA 360
45 GGGGCTGCTG TTGTCTTTGG AGCATGCTGC TACTATGCT TGGGACTGTC TATATAGATT 420
GAGGCTATG AAGAGGCTGT AATTTGGCTT CAGTATGTCA AGATAGAT TCAATTCACC 480
TATATGACT TACAGGAGG TATGTGTTTA ACGCTTTCT CTGCGTAGC AATCGACAGA 540
50 AGCGCTGTC TCAATGACTT CATGATGAGA GCGCTTTGGG TGAATATGCG TGTGACCTTT 600
GAGCGCTGCG TTGAGCTGCG AATGCTGTGA GATCATATAC CATATGACCA GAGCGCGGCG 660
CGAAAGCAT TTGCTTGTGTT GCTACATCTT GGTGTGATGCG GTGAGTGTG CTGCTCTCTG 720
55 ACGATATAG GGGTCTCTCT TCTCATGAGA GCTGCAATGT ACGAGCTGCG CATTTGTGGA 780
GCGCTCTGCA CTGTGCGCAT GTGTGCGCC AGTGAAAGAT TTCTGAACAT GCGTGCACCC 840
60 CTGGGAGTGG GCGTGGGCTCT GGTCTTTGCG TGTCTATGCG GATCTATGTT TCTTCCACT 900

5 ACCACCTGCG CTGGTCCGAC TCTTTACTCA GTGGCAATGT AGGGTGGATT AGTCTTTTTC 960
AGCATGTGTC TTCTGTATATA TACCCAGAAA GTATATCAAG GTGGCAGAGT ATCCACCATG 1020
TATGGAGTTC AAAAAATAGA TCCCATATAC TCGATGCTGA GTATCTACAT GCAATACATA 1080
AATATATTA TGGAGATTGC AACTATGCTG GCACTGGAG GCAACAGAAA GAATGAGAGT 1140
10 GACTCAGCTT CTGGCTTCTC TGTACATCA AATATCTTGT TTATGCGGCC AGATATGCAAT 1200
TAATATATTT GTACAGGAG CTTTGTTTGA AGTTTAGAG ATAGAAAGAA TGTCTATATA 1260
TTTAAATGTT CCGTAAATGT GATGCTTCAG GTTGGCTTTT TTTTCTGGAG AATATATCCA 1320
15 GTATCTCTCT CCAATATAG CACACATAT TTCAATTTCT ATGTTTGTAGT GAATTTTAAA 1380
TGTTTTGGTG AATGTGAAA CTAAAGTTTG TGTCTATGCA ATGTAAATCT TTTTCTTACT 1440
20 TTAATATTTA GTAGCTTCAK TGAATACTA AATTTTAGCA AACCTGTGT TGTATATTTT 1500
TTTGGAGTGC AGATATATGT AATTAATGTC ATAGTGAT TTGGAGCTTTG GTAAAGGAGC 1560
CAGAGAGAG GAGTCACTCG CAGTCTTTTG TTTTATTTAA TACTTTAGAC TTAGCACTTG 1620
25 TGTATATGAT TATGTAGGAG CAGTATAGA AATCTGGGT ATTTGGAAAC AATGTGTGAT 1680
TGTATATCT ATCTGCTGAA CTTAAGAAA CTGTTTCACT TGAACAGAGC AAGAGTATG 1740
30 CATCTCTG TGTTGCTCTC TGAATGCTCT CTTCCTAATA TATATGTGT CATGTTTCAK 1800
TTGTACAGAA TGTATATCAT ACAGAGATC CTTCATGAGAA TTATATATGT GTGTTTACT 1860
TTTGATATGT ACAAAGGAA ATACTTTTA ACTATCTCT AAGAGAAAT ATTCAGACA 1920
35 TGAATATAT TGTCTTTTTC AATATCAA CAGTATATCT ATGATTTCTA AGTGTTTTAT 1980
TATTTTGTCA TATTTATGA ACTGCTAAT TGAATACAGC TTGCTCTTGT CAGCTCTTCA 2040
40 AGCTTTGAG CCTTTATAGA AAGCTCTCT TGTGGCTTAC ACTGGAAAT ATGAAAGGAG 2100
TTTCTCTCT AAGACTTTTG GTTCTGCEA TTGCTCTCA GACTATGAGC TAAAGAGCA 2160
45 AGCAAAACAG AACTATGACT GTCTTAATGA AATATATCAA CCAAAAGTG TAATAGGAA 2220
AATGCTTCAAT TATGTTTCCC TACAGAGCTT TTACTTCTCT TACATCTCTA CACATTAAT 2280
TTCTTGAGAC ATTGTATAGT CCTTTGATAC AGAAGAGTTA TATTTAGGAG GATTTATG 2340
50 AAGGG 2345

(1) INFORMATION FOR SEQ ID NO: 302:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2369 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

5 TTTTATTTT TTTTATTTT TTTTTCAG ATCATGTAT ATTTATTAAT TCGATATAAA 60
AGATATATA CATATTAAGG AATCTCTTAA AATTCAGCTC TAGAGTTATA CAGATCTAG 120
TACTTTTCCA ATGATATGTTA ACACAGCAA AAAAAATCTC TAAACACCTG AAGCCCCAC 180
10 TTTTACAGTG GACTATGCTA ATAAAAAATT TTGACATTTA ATTTGTCTCA CATATATAT 240
TTACATATG AATCAATGCG TGAATATACA ATAAATGAT AAGAAATAG TAAATATAAA 300
15 CTTTAAAGAG CAAAGTTTAA TATGCTGACA ATGCTAATTA TCTTAATTTT ATATATAAA 360
TTAAATACATA GAGCTTTCTG TTACAAATTT CTAAATGCTC TGGTTTATTA TCAATTAATG 420
20 CTACCAATTT ACATGACACA TCTGCTAGCA CTGCAATTTG ATTTTATTTTC CCAATATGT 480
GTGAGTATAT AATGACATTT TCAAGACAGA TATTAATTTTA CTGTGAGACA GAAAAGAAA 540
CTCAGATATG GTACTGTCTA CAAAGCTCTT CCAATATGAA ATTAATATAAA CAGTAAAGTA 600
25 AATTTTAAAA AATATCTTAA AAGGGATGCT ATAGGCAAG AATACCAATTA ATGECACAGC 660
TCAAAAATCT CAGAGACAAA TCAAGACACAC ATCTTTTCTC TCTGCTTTGAG CAGAGAGAG 720
30 TCGATTTTTC CATCAATATA CCAATATTTCA AGCAAGCGAG GGGCAGCAGG TGTATAGCTG 780
ATTAGATCTT GCAAAATACT AAGATGGGAG CAGGGGTGGC CAGAGAGAGG GGTAAATTAAT 840
ATATATTTCA AACTATATAC ACAATATATG GAATGCAAGC CATGCCAATC TGCTCTGTG 900
35 AATCAATTTG ACCTTTATAG TTAATATTAAT AACAGTGTGA ATATATCAAT AGATTTAAT 960
GGCAAGCAAA ATCCAGGCA CATTATTAAT TATATATTTA CTGTGCTGTT TCAATTTAAA 1020
40 AATATTTTGG CTATATATAC ATCTCAAGCTG AAGTCTATAT AAAAAATGTC CTATATAGTA 1080
CAGATATTTA CTTTGTGCTA GTTCAGAGCC TTTTGTGTCAC TTCTGTCTGTA ACTGTAGCCA 1140
1200 GAATCTTACA TGTACATGCA CATATGAGAG AACTCAAGCT GAGGTCAATCC AAAAACTGTG 1260
45 CGTATAGGGA GCGTGGAGGT ACTTTGAAAG TCAATATAGA CCAAGAACCC AATACAGTTA 1320
ACATGTAGGA TGGCAAGAG GAATGGAATG CCAATATGCG AATTAATGCTT TTTTATAAAA 1380
50 CAGAAAGAG AAGGCTCTC GTACAGGAG AATCTGTATC ACATACAAA AAGAAAGAGC 1440
CACCCACCAAT TTTTATTAAC AGAAGCAAT TATATATGTTG GAATATGAA ATTACAGAAA 1500
ACCAAGATCT AAGAGAGAA AATCTACTGG TTTACTTGGG AGAAGGAGA ATGTTTCAAC 1560
55 CCAAGAGAG TTACTTGTGG AAGCCGCGCA CCAAGCCGCA CAGAACCTCA TTGTGTGTGG 1620
CCTTCAAGCA TTCACTTCTA GGTCTTAAGT CAGAAATGTT CCGCACTCTC TTTGTAGCCA 1680
AATCATATG CTGTTCAGA AGAGAGCAA AAGCATTTCT CAGGACGTTCC GAGCATGAG 1680

5 CAGGTAAAT GAGGCGAGC AGGCGCTGT CATTGCGTG AGGCTCATTC ACCATTGT 1740
CAGACGCG TTCTCTACT TTCTTGATCA GCGCTGCTT AATGTGTATA TTGTGAGG 1800
GATGTGTGT CATTGCTAAA AATAGGAAT TCTGTTCCTC TGTGTCAAT ACAGCTTTT 1860
CGACAGGT TTATGCTAAT GTTCCGCTTA CATTTCCTTA CTGTAATGC AATTAAATG 1920
GATTCAGT CTACAGTA AATTAATCAA TCGATTCTG GACCTTTCT GAGGCTGAG 1980
TTTCTTAAC ATGCTTCAGA GTTTCATCAA GAGACATC CCTGTGTGA GATCTGACT 2040
TACAGTAC CTCTGTGT AATGACTTT TAGTGTCAAT TCCAGGCC TGTATGTGA 2100
AGCTTCTCT CATTCTAAT TCAATTACA TAGAGCAGG TAATCCAGT GATTAACGT 2160
CATTCGAAA TCAATGTATA ACTTTCGCG TCTTGAAGC CCAGCAGAG CACTTCTCC 2220
ATCAGGTTCA GCGCGTTTC CTTCGATGCG CCTTGTGT GTTGTCTG CTGTCGCGG 2280
GCGCTTCGG CTTGTCTCT GCTCTAGGC GCGCGCGGC CCGCGCGCG CTTCTGTGG 2340
GCGCGTTTC GCGAGGCTC GATCGCGCG 2369

25

(2) INFORMATION FOR SEQ ID NO: 303:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1181 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

535 GCGAGCTGT GTTTCAGTC GTGCGCTCC CGTGGGTTT GCGAGCTTTA GCGACTATTG 60
GCGCTCGCC AGCGCGGCTG CAGACTGCG GCGTGTGTTG CTGTTCGCG GTATGCTAG 120
GCGCTGCTT GCGCTCAGG CTGTTCGCG GTGTAGGTGG GATTCAGCA CGTTCCGCG 180
CCGAGGTTT CCGGAGGT GCGGCAATG GCGGCAAGG GAGAGATGG CTACGCGAT 240
GCTCTGGTG GACTGAGA TCAAGAGTT GGCATTTGAG AAGCACAGA TTATTCAGT 300
GCGCTCTTG ATATGACT CTGATCTCA CATTGTGGT GAGGTCTTA ACTGATTTT 360
AAGACAGCA GATGATTCC TCGAGCTCA GTATGTTGG TGTAGGAGC ATCAGCGGA 420
GTCTGGCTT ACCAGGCG TGAAGAGAG TACAAATACA TTCCAGAGG CAGATATGA 480
ATTCTGTCC TTGTACAGC AGCAGCTCC TCGAGGCTC TGTCACTTG CAGGAAATTC 540
AGTTCATGA GATTAAGAT TTCTTGACA ATACATGCC CAGTTATCA AACATCTCA 600
TTTATGATA ATTATGTGA GACTGTTAA AGACTGTGC AGAGCTGTG ATCCAGAGA 660
ATATGAATTT GCACCAAGA AGGCTCTTC TCATAGGCA CTGTATGCA TTATGAAAG 720

60

5 CATTAGAG CTTGATTTT ACCGAAATA CATTCTCAG AAAAAATAG ATGAAAGCA 780
GAGAAATTT ATGAAATG GGGAAATCA GAGACGCTG AGTTGATGCC AGTTATCATG 840
CTCCACTAC ATCTATAT CTGAGCACT TGTGTGTTT TTTTTCCTC AGCTGATGG 900
CTTGGAGAG CACTTGGTT AACTTCATC TCGAATGTA TTACTCAGC AGACAGACA 960
CGAAATCTA TTTTCTCT ATATGCTGT TTCCATTAAT ACACAGCAGC TCTTTGTAA 1020
GTACAGCTC ATGTCCATCC CTGTGTACAT ATATGCATTT GCTTTTAAAC CATTTCTTT 1080
GTTTAAATA ATATTAAT AATTAAGCT AGTTCTATTG AATTCGAAA AAAAAAATA 1140
AAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA N 1181

20

(2) INFORMATION FOR SEQ ID NO: 304:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1537 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

30 CTTTGTGT TTCCGCGCAT CCCACTCTC CTCGACGCTG GACGCTACG TTCCGAGGC 60
CGAAATCTG CCCACTCTCC GCGCGGGCTT AGCGGGGTT TCGAGCAGCG GAGCCTTCA 120
GCGCATGCC AACTACAGCG GCGCGCGCGG GCGCGCGCG AATATGAGC TGGCGCGGA 180
TGGGAGGTT TCGAAGAAA CATCGAGGCG GCGAGCTCAG CTGTATTTGA CATGAGAGC 240
ATGATGATA CCTCAGGCT TACTTTCAG GATATGGTG AGCTGCATCA GCGCTGCGC 300
GAGAGAGAG TAGAGCTGA TCGAGCTGAT GCGAGCTCTG CTGAAAGAGA GATTCGAGAG 360
TTCTGGGCA TGAAGGCTT TAGGGAGAG CTGAGCGGCG AGGTGGCAGA TCAATGTGG 420
CAGGCTGGA AAGACAGC CTCAGGGGCG TTGAGCTTGT AGCCAAAT CCAATCTCTC 480
AGACCTACT TTGATGTGA GCTCTCTCAG GTGCGAGAG GGTCTCTGGA GTCCATGATC 540
CCTATGAGA TGTTCAGCTT CCCCAGAAA ATTCAGGCTG AACTCTATGG ACCTCTCATG 600
CTGTCTTCA CTCTGTGTG TATCTACTC CATGGATGA AGAGCTCTGA CACTATATC 660
CGGAGGACA CCTGATGAG CAGAGCAAT GCGAGCTCT TCGGCTACTG GCTGGAGTTC 720
TCATCTTCA TTTCATCTCT TCGCTACTG TCGAGGCGC AGATCACTAT GCTCGAGATG 780
TTGCACTGC TGGGCTATGG CCTCTTTGG CATTCATTTG TCTCTTTCAT CACTATAT 840
ATCCACTCC AGGCTCTTT CTACTCTTC TGGCTTTGG TGGGTGACT GTCCACTCTG 900

CCGATGCTAG CAGTGTGCTGT GTCTGGGACC GTGGGGCCCA CAGAGGGGCT GCTCTCTGT 960
GGCACCCTGG GTGCGCTACA CATGCTCTTC GTGCTCTATC TCGATTTTTC GTACACAAA 1020
5 GTGTATGAGG GATCTGTGGA CACACTGGAG GGGCCGACCA TCCCGCCCAT CCGAGGGGTC 1080
CCGAGAGCA TCGCTGCAAT GCTGCTGCTGT GCTGGGCTTC CCACACGCTT CCGTAAACCC 1140
ACAGCCAAAG CTGTGCGGCT GACCTGTGAG TCACACTGAC CCACCTGAAA ATCTGTGGCC 1200
10 AGTCTCTCTT CCCTGAGCTG CAGAGAGGAG GAGACTATTT AAGGACAGT CCGTATGACA 1260
TGTTTCTGAG ATGGGTTTGG CAGCTGTGAC TGAGCTGTAG GTGCTGTAGT ACCTCTTGTG 1320
15 AGCTGTGGCC ACTTCTGAAA GGCAGAGGCC AGAGCTGCTT GGCAGGACT GCAAGGCTCT 1380
GGAGCCAAAG CAGAAATGAG GTGAGCTGCT TTGAGAAACC GTCCCACTT ACCCTTCTCT 1440
TCTCTTTTAT GTCTCCGACA TTCTCTTGTCT AATATATAGC TTGCTAATTA AAAAAAAAAA 1500
20 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GGGGCC 1537

(2) INFORMATION FOR SEQ ID NO: 305:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1493 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

TCGATCCCAA AACGATGCC TCCCAACAAA ATCTATAGAC ATCCCATAT ATATATGTTG 60
TTATATTTCT ATTGCATCA TTATTGAAA TACCCAGCTC AGTCCCTGGC TTATATATATG 120
40 TTATATGCC TTACTACTC TTGCTCTATTT TTTTATTTTG AATGCGAT GAGCGAATTA 180
ACACATTCAT GGTGAGGCA ATTTTGTGA CATTCTCTGT TACCAGAGA TCTATATATCA 240
GGATGATCTT GAGCTGTCA ACGAGCTGT ATATATACAG ACGATGAAAC TCTTTCCGGA 300
45 GCTGGAAATTT AAGGGAAG AGAGAGAGC CAAGAGATG CATGAGAGA AACGACAAAG 360
GGAGAGAG ATTCAGACTC AAGAAAAGC CAAACCGGAA AGAGATGGC AGAATATCTT 420
50 TGAGAAAGT CGAGATGTC GTGTGGAGAG GTGGGAAAC TTCCAGGCA ATACAGAGG 480
GAGAAAGAG AAGAAATTC GAGCTTCTT GAGCCACCG AAGTAAAAA TGGAGCAGG 540
TGAGTAGGAG CCAAGSTCA CAGGCAAGA AGCTTTGCCC TCGTATCTCC CTCTCTGCTT 600
55 CAGAGACTC ATTCTTCTT CCGACTTCA CCCGACATA GAGTATATTT TCGTTTATG 660
TCCATTTTGT TTTCATATAG ATTATATATC GATCAGAGTA ATCTTTTCTT ACATGAAAT 720
60 GAGGGGCTTG GTTTAAAAA AGAGCTTTTC CTCTCCCTGC CCGTAGAGCA ACCAGTATTA 780

GAGGTGCA CAGTGTGTC TCGTCTCTT TCCACAGCC TGTAACTCAG TGTTTTATAC 840
5 TTCACTGAT TTGATGTTT AGAACTCTG TCGATGATTT GTGAAATCA TCCATATAAA 900
CTACTCTT AAGAGATGT TCTGTGACT TCAGAGACA GCTCGAAGG GCGACCTTAG 960
GAGGCCCCC CCGTTCAGTT GCTGCTCTCT GGTGTGCTC CTTTCGAGG CCGAGATAG 1020
10 ACAGGAGCA CTGTGAGCA CAGAGCAG CAGTGTATC CTTGTGATGT TTGCGATGTG 1080
CCCCCTGCT ACTGAGCAAT CAGTGTGCA TGAAGGCCAC GCGACCCAAA CTTTCACTT 1140
TCCAAAGAG TAGCCCTCT CCAACCAATA CAGTGTCTA GCTGTCTGC ATTTGTATGT 1200
15 GGTATATTC TTATGTATA ATAAATTTT ATACCGAGC CATTGATGA CTTTCTCTTG 1260
TACTCTCTT TGTGGTCCC TTGTCTGCT TGTCTGAGC CCAAAATCT TTGGGTTGG 1320
20 ACAGACTGG CTGAACCTTA GTTCTCTCAT CTATGAAATG GGAATATGAA TTAAGTCCAG 1380
AGCTTTAGG CGAGTTTGC CATGCCATAT AAGAGTAAC TACATATGT CTCTTGTGT 1440
25 ATTCCATA TCTATATTT TCTGTGAAA ATGAGATAC TGATTTTGT GAG 1493

(2) INFORMATION FOR SEQ ID NO: 306:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 577 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

AATTCGAG AGGATATATA TACACTATAC TCGCATTTAC TGTTCACCC AGCCCGGAAA 60
GTGAGATG TATATGGAA AATTAGAAC TCACTATACA TTGTTTCCA GAGGCTCTC 120
ATAGCATTT ACCGAGAT CTACAGCAT GATTAAGACA CTTATATTCG TTATGAACTT 180
45 GACTATATC TATATTTTTA TTTTATTTT TGTGTTTAT GCGAGCTAC TTCACTCTT 240
AAGCTGCTT TCAATTTGT ATGTAAACT TTAAACATG CAGATCAGTG TAGAACTGT 300
CATAGAGAA GAGTAGAAA TCGAGTACA TGATTTTAA ATACCTCTTC TTGTTTTGG 360
50 ATGTTAAAC GTAAATGCA GTAGTAGCA AAGACAGAT GATTATATAC ACTATACTG 420
AGGATTTCA TTTTAAATC ATCTTTATGA AGATTTAGAA CTCATCTCTT GTGTTTAAAG 480
55 GGAATGTTA ATTAGAAT AAGATTTGT GAAAGAAATG TTAAGAAAA AAAAAAANA 540
AAAAAAAAA AAAAAAAAAA AAAAAAAA AACTGCA 577

(2) INFORMATION FOR SEQ ID NO: 307:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2660 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

5 GTTGTGACCG CTCCTGCAAT ATGGCTCCCG CGGGCTGCGA GRRKTCRGT CNCRGTGGC 60
15 TAGAGCTGTCC TGAAGAGGGA GAATTAAGCT CCGTTCTCC ACCGTGCGG CTGGCCAGGT 120
GGGCTGAGGG TGAAGGAGAG ACGAGACTT GCTTGTGGA GCTTAATGCT CAGAGCTGG 180
GAGGAGGTT CCGCGCTTC TGTGCTGTCA GCGCGGCGAG CCGCTTCGCG CTTCATTTCC 240
20 TCCCGCAGCC CCGCTACTG AGAAGCTCG GATTCCTGAG AGCGCGCAG CCGTGGCTTC 300
AGCTTCGCG GCTTCGATC AGCGCAGAC CAGCGCGCAC TGGGAGGGA CAGAGAGCC 360
TTGACATCTC CAGCTGACCA GAGGTGCTCG CTGCAACCGA CGAGCTCTCT CCGTCTAGGA 420
25 TGACTTGACC CTGAGCTCT CAGTTTCTCA GCTTGGAGAC ATTGATGGA GCGCGAGAG 480
ATGGCTCTGA GCGGAGCAGA GAAGGTGCTG ATTGTGGAG CCGGCTGCTT CCGATGAGT 540
30 CAGAGTTTCA GCGGAGAGAC GCGAATTTG CCGCTTCTGA TAGAGTTTCA CCGTCAGCTA 600
CCGAAAGGGA AGAGTTGCTA GTACGCGGGA TCGAAGCGGA TTGAGCGGAG ATCGGCTTT 660
35 CAGTGGTTC TCCCGGGTG TCCCGGAGA TGTGGCTGGA CTTTCAGAGT ACTGAGCA 720
GAGCAGGAG TACCTGACCG ACTTGGGAA TAGAGAGAG GCTTCAGAG GTAGAGCGCC 780
TATGAGGCG TGTCTGAAA CTTTAGAGAC GGGTCAATG CTTGCAATCT GCGACTGCTG 840
40 CAGATCGACC GGGACTCTCG CAGTCTGAG CCGCTTGTAA ATGCCGAGTG CAGAGATGAC 900
TATTAAGGAG GCGAGAGCC TGTGCAATC CCGATTGGA AGAGAGATC TCGAGTGTGT 960
45 GAGCTCTCT GTGGAGATG GCGCAATGT GATGCGCGG GTCTGCGGCG ACTTCTTCCA 1020
GAGCGCGCA GCGACTTCTT TTTATTTGCG TAGCTTACCC CTCTCTTTG CCGCTTGCAC 1080
CAGAGATGCG GATGTGTAA GCTTACCTCT GAGAGACCCA CACAGCGCG CCGAGCTTGA 1140
50 GCGCTGAC TCCGAGGGA AGAGATGCT GATGCGCTA GTGATGATC TCGGAGACT 1200
CAGCTGAGAA CAGTGCATG GTGACAGCA TGTATGATG GCTTCTTCAA GCTTGGGCG 1260
55 GCGTCTGCG CTACGCTGCA GCTTGGAGAC ATCGGAGCC TGGAGATCT CAGCGCTG 1320
AGCTGCGCG CCGAGGAGCG CAGATTCAG ATTTCAGCG ACATCTTCCA GCGGAGTTT 1380
TGAAGATCA GCGACTTTC CCGAAGTTC ACCGATGCT GCTATGCGCG TGTGCGGCTG 1440
60 TCGCTGATG AGCTGGCTTC TGTGAGAGC TGTGAGAGA ACTGATGCT GAGATCATT 1500

CGCTTTCATT GCAAGAGCC GCAAGAGCAC CGAATGCTCG TTTTGGAGCC CCGAAGCAA 1560
5 CTCTGCGAG CGAATGGA TCTGCTCATC CCGAGTTCT TCTTAACATT CTTGTATAT 1620
CTGATTCACA TTTTCAATCT CACGCTGTTT GCTTACCATC AGCTTACCTT CAGAGAGAG 1680
CGCGCGCTTC ACTGAAAGC GAGGTTTGGA AACTGCAATC TGTGAGCGG CAGATCTTT 1740
10 ATCTGCTGAT GGGGATCTA CTTCTCTGTC GGGCAGCTG TGTACTTCT GCGCGCGCA 1800
CGTGTTCATC TGAATCTCT TCAAGAGAG CTACTTTTGA AATCTCTTC CTTTTCAGG 1860
CGCTCTTCA CAGTGTGTC CCAAGTCTG TTTTCTCTG GCAATGAGT GGTACTGTC 1920
15 CTTGCTTTG TCTGCGCTCG TGGTGGCTT GGTGAGACT GCTTTACTTA TACAGTTGTC 1980
GTTCAGCAC ACAGCAGTC TACAGTTTCA TGTTCCTGTA AGCGCTGCTG AGCTGAGCC 2040
20 AGAGGCTTG GCGCGCGGAA GCTCTTACAG GCGCGAATTC CAGAGTTTCA GTGAGCGCA 2100
TGAAGGAGCA GAGAGAGAG GCGAAGCGG CCAATACAG GGTATCTG GAGGCTCTCT 2160
25 TCGAGCTTT CAAATTCACC ATCGGCTGTC GCGAGCTGTC CTTTCAGAG CAGCTGACT 2220
TCCGCGCAT GGTGCTCTG CTGCTGCTG CTTACGCTCT GCTACCTTAC ATCTGCTGTC 2280
TCAAGTCTT CATGCGCTTC ATGAGCGGA CTTGCAATG TCGGCTTGA CAGCTGAGC 2340
30 ATCTGAGAG TCGAAGAGC CACTCTCTTC CTGAGATG AGAATGCTA TTTGTGCTGTC 2400
AGGAAAGAG AGCGCGGAG TGTATCTG ACCTTTGCA CTAAAGCGAG ATGCGAGCC 2460
CGATGAGCC TGGTCTTCA GGTGAGGGA GGTCAACTG GCTTCAATG CAGCAGAGC 2520
35 TCGTACGCT GTGTAGGAC CCGTACGGG CAGGTGTGTC TCGACTCTC GAGAGCGCTG 2580
TCTGCGCTTC CCGTCCGAG GAGATGAGG ATGGTGGCTC TGAAGAAAC TATGTGCGCG 2640
40 TCGAGCTCT CAGTTCAC TGTGCGCCA GATCGAGAG GAGCGCGAG CAGCAGAGAG 2700
AGGATCTTTC CAGCAGATC TGTGCTCTT GGGTTCGAG TGAATCTG TGGCAATAT 2760
45 ATATTTTCA TACTGAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GCGCGCGCT 2820
ASCGAATTC GCGCTTATG TGAATGCGTA TTACGATTA 2860

(2) INFORMATION FOR SEQ ID NO: 308:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 876 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

CTGCTGTGTCT CTGCGCTGCTCTG CTGAGCTCTGCT TTTTACTATPAC AGCTGGCTTTC 60
CAGCAGACAG GCATCTATGAG TGTATCATGTC CAGAGAGCTCT GCTGAGCTCTG AGCAGAGANN 120
5 TTGGCGCCCC GAGAGCTCTTA CAGGCGCCCAA TGGCAGACAG TCAATGACAG CCAATGAGAGG 180
ACAGAGAGAC GAGGCGAAGG GGGGCTCACTA CAGGCGTATC CTGGAGAGCT CATTGAGACT 240
CTTCAMATTC ACATCTGCTA TGGGCGAGCT GGGCTTTCAG GAGCAGCTGC ACTTTCGCGG 300
CATGCTGCTG CTGCTGCTCTG TGGCTTACTG GCTGCTGCTAC TACATCTCTGC TGTCTACAT 360
GCTCATCTGCC CTATCATGAGC GAGACCGGCA ACHATGCTGCT CACTGACAGC TGGAGCATCT 420
15 GGAAGCTGCA GAAAGCTCAT TCTGTCTCTG AGATGAGGAA TGGCTATCTG TGGTGCAGGA 480
AAGAGAGAG GGCAGGCTGTC ATGCTGACCG TTGGCACTAA GGCAGATGCG AGCGCCGATG 540
AGCGCTGCTG CTTCAGGCTG GAGAGGCTGA ACTGCGCTTC ATGGAGGAGC AGCTGCTGCTA 600
GCTCTGTGTA GAGAGCTGTA GGGGAGGCTG TCCCTCTGAACT TCTGAGAGAC CTTCTCTCTG 660
CTTCCCTCTC CAGAGAGAT GAGATGCTG CCTCTGAGGA AACATATCTG CCGCTCTGAC 720
TCTCTGAGTC CACTGATGAG CCGAGATGCA GCGAGAGGCG AGAGGACAGA CCGAGAGATC 780
TTTTCAGCA CACTGCTCTG CTCTGCGCTC CCACTGAACT CTGCTGCGCA ATATATATTT 840
30 TCACTAATTT AAAAAAAAAA AAAAAAAAAA ACTGGA 876

(2) INFORMATION FOR SEQ ID NO: 309:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2025 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

CATGACCCCG CTGATGCGAT CCGGACAGC CTCTGCTTCC AGCTCATCTT CTCTGATGAG 60
45 CAGCCGCGAGC GCTTCCGCA CAGCGAGAG CAGCCGAGAG CCGCTCCGCA CCGAGAGAGG 120
CAGCCGAGC GCTTCCGCA CAGCGAGAG GCGCCAGCTG GAGATCAGC CCAATCTGCG 180
50 TGTCTGCTGG AACAGCCGCG GCGCAAGTCC ATGAGAGTCT CCTGTAAGC GCGCTGCGCA 240
GCTGCGCCCG CCGAGCTCTG ATCTCTGTAG TGGCGCCCTC CTCCCGCGCC CATTTTGCGC 300
CCCTGCTGTC CATTACTGCG CTAACTGCTT ATTAACTCAA AGCTTATTTT GTAGAGGTGA 360
GCTCTGCTGG AGAAGATGTA GGTCTATTTAC GTGGGTGCGC TTCTCAAGG CCGGGTGGCG 420
GTGGACAAA GGAAGGAGC AAGCATCTCC GCATGCGATC CTCTTCATTT AACCATGCGC 480
60 GGGTTGACAC TCTCTCTCCC TCTCTGAGAG AACCAAACTT GCGAAAAACA AGACGGTAC 540

AGCAGACACT TCACAGAGCC AGGCTAGGC CGCCTGAGC ATCTGTGTTT AACGGGTGTC 600
5 CTGGTCAGAA GGGCAGCGCC CCACTTCCCG TTCTCTCTTT AACGAGGAG AGCTGATCC 660
AGTTTCGCGA AACAAATCC TTTTCTCATT TGGGAGGGG GGTAAATGTC ACATGAGAGC 720
ACCTCTTTTA AACAGCGAAA ACAGAGAGG GCGAAAGGTG GATATCATCT CAGGCTAGA 780
10 GGCATTTGGA ACACAAATC TACGTAGTTA ACTTGAAGAA AGCGATTTTT AAGTTGTGTC 840
CATCTAGAAA GCTTTGAATG CAGAAAGCAA CAGCTTGAT TTTCTAGCA TCGCTTAAAT 900
GTGAGCGAAA AGCAGGCTAC AAATCTCTCT GCTTTTACAG ACGAAATATTT TTGCGAAGC 960
15 GTTGGGCAAT ATGGTTTTTG AAGGCTTTAG TTCTGCTTTC TGGCTCTCTT CACAGAGGCC 1020
AAGCTCTCAC CCGTATPACA TGAAGCATG ATTAATCTTG TTGAGGAGAA AGATCATTTA 1080
20 GATTTGTTTT GCAATCTTTA GAATGAGGG CACATATCCA CAGCTGCGCT GCGTGTGATG 1140
AGTGTCTCTG CAGGCGCCCG AGTAGAGCA CTGGGCTGGG GCGGAAATG GGGTTACTCG 1200
ATGTAAAGGA TTCTCTGTTG TTGTGTGAG ATCCATGCA GTTGTGATTT CTGTGATCC 1260
25 CAGCTTGGTT CAGGAAATTT TGTGTATGTC GGTAAATGCC ATTTTCAAT CTTCAGACAG 1320
TGGGCTGAAA CGTGAACTCA GTAGCTGAAC CTGTCTGACC GGTCACTTT CTTCGATCTT 1380
30 CAGAACTCTT TGGCTTGTTC GGGTGGGGG TGGAACTCA CTTGGGAGC GGTGCTGAG 1440
AAATGTAG GATTTCTGAA TACATATACC ATGGCACTTT CTTTCCCTCT CTTGCTTCTT 1500
35 CTTTTCCTGC TCGCTAACTT TTCCCGAAT GGGGAGAGC CACTGAGCTT TCTGGCGGCG 1560
CAGTCCGCTT GCGAGTTTCC TGTACTACTG CTTTGTACTT TTCAATTTTG CTGACCTGCG 1620
ATTTTCTCAT AGGAAGTTTG GTGAGAGTGA ATTGAAATTT GTAACTGAGC CACTGGAGCC 1680
40 CAGGATTTTC TGGAGCCCG CAGTTGGGAG GAGGAAGTAG TCCAGGCTTC CAGGTGGCTT 1740
GAGAGGCAAT GACTGTTTAC CTGCGCCCA TCAGCTTGA GGGTTTCTCT GCGCTTGAAT 1800
45 AAAAAATGTC GGAATCGGG CAGAGAGGC TGAATTAAGAA TGGAAACTA TTGTGACAAA 1860
GTCTTTTCAG AGAGTTTCTT TATGAGATA TTGTATTTTA TTTCAGAGC AATTAATTTG 1920
TAATTTTGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1980
50 GAGGCGGCGC CTTAACCAAT TCGCTTATA TGAATGTAAT GATC 2025

(2) INFORMATION FOR SEQ ID NO: 310:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3026 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

5 TAGGAGAGAC TGAATATCC TAAACCCCTA AGCTCCAGGT GCCCTGTGGA AGGAGCAACT 60
GAGTATAGC AGGCTGAGGC TGTGTCTGCC TGGTCAATAG CTCACCTCTTT CCCCCAAATC 120
TTCCCTGGA GCTTTGAGAC CAAGGTGCTA AAAGGAATAG GTAGAGAAC TCTTCTATCT 180
AATCTTAAA ACCGATATGT TGAATCTCA TTCAACAGCT GATGCCCTAT AACCCCTGCC 240
TGAATTTCTT CCTATTAGC TATAGAGAT AGCAAGATCT TTACATTAAT CAGAGTGGTT 300
TCAATGGCTT CCTAACCTCT CTATAGGCC CTCCTATTAT TTGACTATAG CATTACAG 360
TGGACTAGC AATTATCAA GATATATGAA AATACAGTGC TTATAGGCTC TAACTATACT 420
GCCCTCATA TGAAGCTGC CTGGGAAG CATTGAGGCC TCAAGGCTTC CTATGTGCT 480
CCACACAG AGCTCTGTA TGAAGTCAAT CTTTTTCCC TATCTCTGTC TTCCCTGCC 540
GCCCTCAAT GGTATAGGCG TACCGAGGCT GGTCTCTGGG CTAGATAGTG GGGACCAAGT 600
TCAATACCTC CCTATAGT CTAGATAGT AAATAGGCT AACATAGCT ACATGTGTA GTGGAGAG 660
CTGGGTCTTC CTATATAGC CACTGCTATC TACTCTTAGC TGGTCAAGCC GCTGCTTCCA 720
GGTATAGGAC CTGTAGTGT TGAATATAC TGAATAGGA GAGGGAATA CAGAGAGGCC 780
CTCTGCTGTT CCGTGCCTA GCGAGCTGCC CACAGGCAAT AAACCAATTA AACAGATTA 840
CTGAGTCAAT TTTTATGTC GGTCTCTCTC ATTCCAGCT CACTTGTGTC TGCTTTGGCT 900
GACTGGAGAC ACCGATAGC TACAGAGCT GACAGAGGA CTGGAGCTG TGCATCTTA 960
GCTGGAACT TACTGTGTA AATAACTTTC AGACTGCTA CCAATAGTG AAATGCGAC 1020
ATTTGCTTT AATATCTTA CCAATGTTGG GAATACTGG CTTTTTCCA GCGCTTTCCA 1080
GGGCAATAAA CTCAGCCCTT TGGATAGGA GTCCGATCAG CCTATATTT TTTTAAAGAA 1140
AACTTCACT TGTTTTCTT TTACAGTGA CTTCTTCTCT GCCCAAAAT TATTAATCTCT 1200
45 AAGGTGAAA AAGATGTTA ACACAGCTT CTTGCTGTA AAATATGTA TTATGACTCT 1260
GTATTTTAA ATTCTGCTC TGAATATCA CTGTCCATCT CTCACCTAC TGCATTTGG 1320
GCCTTTCCA TTGCTGCA TGTCTTAT CATTGAGGC CAGTGGAGC AGGAGAGG 1380
GAGAGAGGG GTCCGACAA CTCGTGTTGC TTTCTAGCT ATCTGAGCA AGAAGAGTGA 1440
1500 AACTAGAGC GCTGCTGCC ATGCACACT CTCGAAACA CTTATCTCC TGCAGAGTG
55 GCTTTGAG GCTCTTACT GGGAGCAAT TAAAGCCCT CTTCAAGCT TCTTTTTC 1560
TTCTTACT CTTTGGCTT CAGAGATTT TGAAGAA ACAATATCT TTACACTAT 1620
TTTCAATTC TAAATTTCA GGGATAGT AAATATAGG CAGTGGCT AAGCTGCTG 1680

TAAAGTTCAG GGGAGAGAA ATCTTAGAT TACAGATTA AAAGCAATC CCTTAACAA 1740
AAAGCAATC AAGACTGCT TCTATTTTG CAGCTTTCC TTTTATGAC AGCTACTAC 1800
CTGGAGAG TAAATTTTCA TTAAACAAAG AAATGGGTC AGCTGAGCTC TGAAGAGCTG 1860
10 AATCTAGCG CCACTCAAT CAGCTTACA GATGCCAGG AGTGTCCAGG AGTTCAGCT 1920
CCTTAACCTG AGCTAGTGA ATAAACCTGG CCAAGTGGAG CAGAGCAAT CAGAGCAAT 1980
CCATCTGTA GGTATAGGC AGCATGAAA GCAAGAGAG GAAAGAGTA TCAAGAGGAG 2040
AAAGAGATC ATTATGTTGG GTCTGAAAGG AAAGTCTTT GCTATCTGAC ATGTACTCT 2100
15 AATACCTGTA AGCAATTTAG GTCCAGATC GCAAAAAAA ATCAGCTAT GGTATATTA 2160
TAAATGCTTT TCCCTGGAGT CAGTTTATTT AAAAGTTAA CTTATAGTTT TTAATGTTT 2220
AATTCATAAA CAGAGAGGAG CTAGAGCAT TCCCTGTAGG AGTAAAGATA AAAGAGTAGG 2280
20 AAAGATTTCA AAGCTCAAT AGATCAGAG CTTTCCAGG TATTAACCT AAATTTAGA 2340
AGTACATTA CAGAGAGTGG AAATGATCT AGTGTCTGAT AGCTACGAC AGAGCAAGT 2400
25 AATTATTAAT TTGAATCCA AACTACTTTC TTATATCAC TTTGCTCTCC ATTTTTCOA 2460
GGAGAGAAA TATGTCCGCC CTTACTTTTC TTGCTTCAA AATTAAATC CAGATTTCCA 2520
AGATCATCT ACAGTAAAT TTGCACAGAC ATCTGCTCAC CCAATGCTCT GTCTGAGCT 2580
30 CAGCAGAGT CAGCCAACT ACTGTGTTGT GAAACCACT GCTTAACTCT TCTGGGGGAG 2640
GGGATTAGC TACAGTAGGA GACCCAGAG TGAATGCGAA AGGTTGAGAA CTTCACAATG 2700
35 TTGGCTGTC AGAGTTGCTG CTAAATTTCTG TCCAGAAAT AGGTTGAGCA CAGCTTTGTG 2760
GGAAAACT GTGGTTTGTG CTAAATTTCTG TCCAGAAAT AGGTTGAGCA CAGCTTTGTG 2820
GGTGTGATG AGGATTTGGG AGCTGTTAT GTTCTTATTC TGGACTGTG AATTTTGTG 2880
40 ATGTAAACA GAAATTTCTG TAAAGCTAT GTCTGTATTA ATATGAGCG TTAAACAGT 2940
AAATATTTCA ATAGAGTCT AAAAAAAA AAAAAAACT CAGAGGGGGG CCGGTATCC 3000
45 AATTTCOA ATAGAGTGT TATTAC 3026

(2) INFORMATION FOR SEQ ID NO: 311:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 712 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

60 CAGGCTTGG TGTCACTTA CAGCTGGGT CAGCAGGCT CAGCAGCT GTTCTCTTT

CTCTCTGTC ACCAGGGCT TTCTGCTGCC GATTTGGGAC TGTGGAMTGG TGTGGGTGCT 120
GTGGTCTGCT CCATGCTGG GTCTGCTGG GTGTGGACT TGTGGCCAA GCACGTGGAA 180
CTGCTGCTCT TGTTGAGTTC GTCTGCTGG TTCTGCTGG GTGGCTTACG CTCTGAGACT 240
GGCTTGCTCT TCACTCTGA CACTCTGGG GCGACGATGG AGCTTGGCAC AATCTTGAGA 300
GGGTGAGCT TGTGAGCT ATGTCTGGAG CACTCTTGG GAGGCTTGT CACGAGATTC 360
ACCTTACG TGATGATGG CTGAGCGAG CTGGCCCGCA GGGCTTGGAG GCGACACT 420
ACAGCTTCT GGCAGCTGG GAGCTCTGG GGAAGTGTCT GTTGGGACT CTGGGAGGC 480
CTGGCTGATG GCTTGGGGC ACATCTCTGG TTCTGCTGG TGTCTATCT CTCTGCTTT 540
CGCTTCTGT ACTTGGACT AGACGAGAG ACTTCTTCT GAGCTGATGG GCTGGATGG 600
TCATTAAGC CACATGTGCC TGTGGCCAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 660
AAGTGGAGG GGGGCGCGGT ACCGATGG CCGATATGA TGTAAACAA TC 712

25

(2) INFORMATION FOR SEQ ID NO: 312:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1289 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

CAAAATTTCA GAATTTGAG GAGGGGAGA GAATATCAA CAAGATTTC TGGAGTATT 60
TTGGCAACT TGTGTTGAG CTGGAGAAA ATATTATGG TGAAGACTTT TCTGTTTCC 120
GTATTTGGT TTTTGGTTG TTTTGGTTG TTTTGTACTA TGTCTTGGTC TGTAAATA 180
TGCAGCTGAA CTACATTCAG AAGGAATAT TGTCTACATA GAATATTTA TGAAGTGGT 240
ACATATTTCT GATGAGAAA AAAAACTTT GCAATCTTT AGCGATTT CTGTTTTTC 300
TGTGTTTTT TCCCTGATG AATATATCG TATTATGAG ACAGCATTT ATTCAATGT 360
TTAGACTTAT TATTTGTTT TGTCTCTGA TTTATGATA TGTGTATTT GGAAGTATT 420
GCGTTTTTA AGCGAAGCTA TATTTGATA CATTAGTAAA AAGGGAATGG TGAACCTTT 480
GTGCTCTTC CACTGAGAT ACACAGAGC ATGTATATCC ATTCTCTTGC ACTTCTTCT 540
TCTTATCTG TTTATACCT TTTATTAAT TTGTAGAGG ACAGGAGTG GCGAAGGGA 600
AGAGAGACT TATTTGACTA ACCAGCGCT CTGTGTTCCA CAGAGCTTT GCTTGTGGG 660
GAGGCTCTC ATTCAGAGG GCGCAGGAG GGAAGAGAA GTGGGCGAAA GCTTGGCTC 720

60

GGCTCTGG AGCTTTGCA TCTGAGCAC GCTCTCTCA GCGCATGCTC CTGAACTTG 780
GAATGTCAA CCGAGCGCT TACAGAGCC CTGAGCATC TATAGACTT GATCTACTC 840
TAAAGGATA TTTATTTCAA CTTCTACTA TTTTACTCA GTCCAGGAC TAACTCTGA 900
ATGGGATTT TCGAGCTTC ACCGAGATG CGAAGCGTTC CCGTGGGGC TGTGGAGCG 960
GGCTATTTCT TCTCTGTTC CAACTTTGCG GTCCGAGCTC CTGCGCGCC ATGCGGTGAC 1020
CGCTCTGTC TCTGTCTCT TCCATCTG TGTGTCTAGC TGAAGAGCA AAGAGAGCC 1080
CGTGGCCCA GTCTGGAGG TCGTGGAGA AGCTCTGAC GTCTCTGAG TCGAGCTTT 1140
GCAATGCAAT TCGTTGTCT GCTTATTC TCGAGATCT GTATAGGCT CCGCTATGA 1200
AATATGCTT CTCTATCTG TATTAAGAG ACTTTTAAA GCAAAAAAA AAAAAAAA 1260
CTTGGGGGG GAGCGGTAC CCAATATC 1289

(2) INFORMATION FOR SEQ ID NO: 313:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Ile Leu Phe Ser Ser Ser 1
1 5 10 15
Leu Pro Phe Leu Trp Leu 20

(2) INFORMATION FOR SEQ ID NO: 314:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 128 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Met Met Phe Leu Thr Gln Gly Pro Leu Pro Ser Thr Arg Ala Arg 1
1 5 10 15
Pro Thr Cys Gln Ala Gly Ala Leu Pro Lys Pro Ser Gly Leu Leu Gly 20
25 30
Val Thr Cys Trp Asn Gly Leu Lys Gly Pro Leu Cys Gly Asn Arg Cys 35
40 45
Ser Pro Asn Thr Leu Leu Ala Ala Arg Gln Ala Leu Trp Lys Gly 50
55 60
Arg Gly Arg Thr His Gln Asp Leu Pro Gly Pro Leu Gln Gly Arg Gln

546

547

65 70 75 80
Leu Gly Pro Glu Pro Lys His Leu Ala Leu Leu Pro Pro Arg Gly Gln 95
5 Glu Ala Ser Trp Ala Ser Ser Leu Pro Gly Gln Gly Pro Leu Pro Leu 110
100 105
Pro His Ile Asn Cys Thr Val Phe Ser Leu Lys Ala Ser Phe Ile Lys 125
115 120

15 (2) INFORMATION FOR SEQ ID NO: 315:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:
20 Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu 15
1 5 10
Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser 25

30 (2) INFORMATION FOR SEQ ID NO: 316:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:
40 Met Asp Gly Phe Ser Ser Arg Leu Phe Ser Ser Leu Pro Phe Val Ala 15
1 5 10
Leu Gln Trp Phe Ile Val Ile Ser His Leu Leu Ser Leu Ser Leu Ser 30
20 25
Ala Cys Cys Tyr Gln Thr His Cys Ser Leu Xaa Gln Leu Ser Ser Ala 45
35 40
Phe Ser Xaa Met Gly Glu Ser Cys Val Gly Glu Arg Glu Tyr Xaa Phe 60
50 55

55 (2) INFORMATION FOR SEQ ID NO: 317:
(1) SEQUENCE CHARACTERISTICS:
60 Met Val Cys Ser Ser Leu Cys Asp Ile Gly Ile Ile Thr Phe Phe

(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:
5 Met Pro Leu Ile Asn Leu Leu Leu Leu Tyr Tyr Val Pro Asn Gly Gly 15
1 5 10
Lys Gln Asp Lys Lys 20

15 (2) INFORMATION FOR SEQ ID NO: 318:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:
20 Met Gly Arg His Leu Val Leu Val Met Phe Ile Thr Thr Ser Leu His 15
1 5 10
Ser Gly Thr Pro Val Pro Glu Asn Val Ile Cys Gly Val Thr Lys Gly 30
20 25
Pro Gln Gly Lys Lys Lys 35

35 (2) INFORMATION FOR SEQ ID NO: 319:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:
40 Met Leu Trp Trp Ser Arg Asp Tyr Thr Met Val Phe Leu Leu Phe Thr 15
1 5 10
Met Val Phe Thr Gly Asp Leu Val Ile Arg Gly Arg Thr Glu Leu Ser 30
20 25
Leu

50 (2) INFORMATION FOR SEQ ID NO: 320:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 88 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:
60 Met Val Cys Ser Ser Leu Cys Asp Ile Gly Ile Ile Thr Phe Phe

1 Ile Val Phe Arg Leu Arg Glu Val Trp Gln Ala Leu Pro Leu Ile Leu 15
20 30
5 Phe Ala Val Leu Gly Leu Ala Ala Gly Val Thr Leu Leu Leu Pro 45
35 40
10 Glu Thr Lys Gly Val Ala Leu Pro Glu Thr Met Lys Asp Ala Glu Asn 60
50 55
Leu Gly Arg Lys Ala Lys Pro Lys Glu Asn Thr Ile Tyr Leu Lys Val 80
65 70 75
15 Gln Thr Ser Glu Pro Ser Gly Thr 85

(2) INFORMATION FOR SEQ ID NO: 321:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

1 Met Gln Pro Gly Ala Gly Val Leu Val Leu Gly Leu Leu Leu Pro Pro 15
10
30 Pro Gln Ser Pro Ser Leu Ser 20

(2) INFORMATION FOR SEQ ID NO: 322:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

1 Met Thr Phe Thr Leu Gly Asp Ser Gln Val Leu Leu Ile Asn Leu Phe 15
5 10
45 Pro Ser Met Pro Ser Gly Ser Cys Ala Arg Pro 20 25
50

(2) INFORMATION FOR SEQ ID NO: 323:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

1 Met Cys Leu Glu Cys Trp Ala Glu Asn Leu Gly Pro His His Thr Ser

1 Ser Leu Leu Asn Pro Arg His Leu Pro Ser Ile Pro Ala Met Phe Pro 15
20 25 30
5 Val Ser Ser Gly Cys Phe Gln Glu Gln Gln Met Asn Lys Ser Leu 45
35 40

10 Val Ser Cys Leu Phe Val Leu His Phe Val Leu His Cys Ile Phe Xaa 60
50 55 60

15

(2) INFORMATION FOR SEQ ID NO: 324:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 196 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

25 Met Leu Ser Thr Ser Glu Tyr Ser Gln Ser Pro Lys Met Glu Ser Leu 15
1 5 10
Ser Ser His Arg Ile Asp Glu Asp Gly Glu Asn Thr Gln Ile Glu Asp 30
20 25
30 Thr Glu Pro Met Ser Pro Val Leu Asn Ser Lys Phe Val Pro Ala Glu 45
35 40
35 Asn Asp Ser Ile Leu Met Asn Pro Ala Gln Asp Gly Glu Val Gln Leu 60
50 55

Ser Gln Asn Asp Asp Lys Thr Lys Gly Asp Asp Thr Asp Thr Arg Asp 80
65 70 75

40 Asp Ile Ser Ile Leu Ala Thr Gly Cys Lys Gly Arg Glu Glu Thr Val 95
85 90

Ala Glu Glu Val Cys Ile Asp Leu Thr Cys Asp Ser Gly Ser Gln Ala 110
100 105 110

45 Val Pro Ser Pro Ala Thr Arg Ser Glu Ala Leu Ser Ser Val Leu Asp 125
115 120 125

50 Gln Glu Glu Ala Met Glu Ile Lys Glu His Pro Glu Glu Gly Ser 140
130 135 140

Ser Gly Ser Glu Val Glu Glu Ile Pro Glu Thr Pro Cys Glu Ser Gln 160
145 150 155

55 Gly Glu Glu Leu Lys Glu Glu Asn Met Glu Ser Val Pro Leu His Leu 175
165 170 175

Ser Leu Thr Glu Thr Gln Ser Gln Gly Leu Cys Leu Arg His Pro 190
180 185 190

60

550

551

Lys Lys Lys Lys
195

- 5 (2) INFORMATION FOR SEQ ID NO: 325:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 252 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:
Met Gly Gly Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg Arg Lys 15
Arg Leu Leu Glu Gln Glu-Lys Ser Leu Ala Gly Trp Ala Leu Val Leu 30
20 Ala Xaa Xaa Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp 45
Phe Gly Gly Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr 60
25 Leu Trp Leu Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val 75
Val Pro Gly Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val 95
Met Gly Val Cys Cys Thr Ala Leu Leu Val Ala Val Val Ala Arg Lys 110
35 Leu Glu Phe Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp 125
Ile Gln Tyr Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln 140
40 Glu Ala Trp Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala 155
Ala Arg Xaa His Gln Arg Xaa Leu Leu Ala Ala Ile Asn Ala Phe Arg 175
Gln Val Arg Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met 190
50 Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn 205
Leu Ser Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu Ala 220
55 Gly Lys Leu Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly Pro 235
Arg Gln Leu Pro Glu Pro Ser Gln Gln Ser Lys Xaa 250

- 5 (2) INFORMATION FOR SEQ ID NO: 326:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:
Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val 15
15 Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala 30
Val Ile Lys Asn Asn Ser His Tyr Gln Thr Ser Lys Ala Leu Glu Leu 45
20 Glu Lys Thr Thr Glu Asn Lys Glu Ser Asn Pro Phe Ile Leu Gln Val 60
Asn Lys Leu Xaa 65
25 (2) INFORMATION FOR SEQ ID NO: 327:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 84 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:
Met Gly Glu Gly Lys Asn Gly Phe Gly Gly Phe Val His Thr Ala Asp 15
40 Ala Cys Trp Glu Gly Val His Ser Glu Pro Val Cys Arg Thr Val His 30
Thr Val His Thr Cys His His Gln Ala Phe Leu Val Leu Ile Gly Trp 45
Ser Lys Ser Gly Lys Glu Arg Lys Glu Ala Phe Leu Thr Ala Ile Ile 60
50 Leu Asn Ser Arg Ser Ile His Ile Ser Cys Ser Trp Pro Pro Ser Pro 75
Val Pro Gln Xaa 80
55 (2) INFORMATION FOR SEQ ID NO: 328:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

5 Met Leu Leu Ile Asn Leu Leu Trp Leu Val Thr Met Ile Lys Ser Val 15
10
Ile Asn Asn Ile Ile Leu Phe Leu Lys Lys Lys Ser Leu Phe Phe 20
25
10 Ile Asp Ser Val 35

(2) INFORMATION FOR SEQ ID NO: 329:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

20 Met Thr Phe Pro Phe Glu Lys Lys Ile Val Ala Phe Ser Ala Phe Tyr 15
25
Leu Ile Pro Gly Glu Ser Arg Leu Leu Ala Pro Thr Phe Asn Pro Ser Ala 30
35
Asp Met Thr Val Ile Leu Arg Gly Arg Ala Gln His Lys Thr Ala Met 45
Leu Glu Ser Tyr Asn Trp Lys Val Ser Cys Gln Leu Arg Glu Xaa 60
50

(2) INFORMATION FOR SEQ ID NO: 330:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

40 Met His Ser Lys Gly Ser Ser Leu Leu Leu Phe Leu Pro Gln Leu Ile 15
45
Leu Ile Leu Pro Val Cys Ala His Leu His Glu Glu Leu Asn Cys Cys 20
25
Phe His Arg 35

(2) INFORMATION FOR SEQ ID NO: 331:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

5 Met Gly Ala Leu Val Leu Leu Cys Leu Leu Val Gly Val Gln Gln 15
10
Ser Gly Ser Val Trp Asp Ser 20

(2) INFORMATION FOR SEQ ID NO: 332:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

20 Met Gln Ser Ala Glu Ile Leu Ser Trp Thr Asp Val Leu His Asp Phe 15
25
Leu Phe Ser Leu Phe Leu Trp Pro Ala Phe Glu Asp Arg Ala Leu Leu 30
Ile Phe Thr Leu Asn Gln Ile Val 40

(2) INFORMATION FOR SEQ ID NO: 333:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

40 Met Gln Ser Leu Val Gln Trp Gly Leu Asp Ser Tyr Asp Tyr Leu Gln 15
45
Asn Ala Pro Pro Gly Phe Phe Pro Arg Leu Gly Val Ile Gly Phe Ala 20
25
Gly Leu Ile Gly Leu Leu Ala Arg Gly Ser Lys Ile Lys Lys Leu 35
40
Val Tyr Pro Pro Gly Phe Met Gly Leu Ala Ala Ser Leu Tyr Tyr Pro 50
55
Gln Gln Ala Ile Val Phe Ala Gln Val Ser Gly Glu Arg Leu Tyr Asp 65
70
Trp Gly Leu Arg Gly Tyr Ile Val Ile Glu Asp Leu Trp Lys Glu Asn 85
90
Phe Gln Lys Pro Gly Asn Val Lys Asn Ser Pro Gly Thr Lys Xaa 100
105
110

(2) INFORMATION FOR SEQ ID NO: 314:

5 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 106 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

10 Met Ala Pro Ser Leu Leu Leu Ala Pro Leu Cys Ser Leu Glu Ala
 1 5 10 15

15 Val Leu Ser Ser Pro Leu Glu Lys Gln Cys Gln Leu Pro Gly Ile Phe
 20 25 30

Cys Gln Leu Gln Leu Pro Cys Pro Leu Leu Leu Ser Ala Gln Leu Leu
 35 40 45

20 Lys Gly Ile Val Xaa Pro Arg Cys Pro Ala Ser Leu Pro Gln Pro Pro
 50 55 60

His Pro Ala Pro Ser Trp His Leu Pro Leu His Cys Thr Glu Arg Xaa
 65 70 75 80

25 Pro His His Leu Pro Leu Gln Gly Gly Ser Ser Asn Met Glu Glu Xaa
 85 90 95

30 Asn Tyr Arg Gly Tyr Xaa Asp Ala Gln Leu
 100 105

(2) INFORMATION FOR SEQ ID NO: 315:

35 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 50 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

40 Met Thr Thr Cys Leu Phe Gly Leu Leu Ser Cys Glu Met Ser Ala Gln
 1 5 10 15

45 Val Ser Gln Lys Ser Cys Val Tyr Asp Glu Ser Glu Cys Phe Ser Ser
 20 25 30

Val Gly Gln Leu Leu Ala Leu Leu Ile Leu Val Tyr Val Leu Pro Ser
 35 40 45

50 Ile Xaa
 50

(2) INFORMATION FOR SEQ ID NO: 316:

55 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 48 amino acids
 (B) TYPE: amino acid

60

(2) INFORMATION FOR SEQ ID NO: 316:

5 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 106 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

10 Met Leu Trp Lys Cys Ser Gln Asn Ile Ala Arg Cys Leu Leu Leu
 1 5 10 15

Leu Ala Leu Val Glu Ile Lys Leu Glu Asp Leu Gln Ser Gln Leu His
 20 25 30

10 Pro Thr Trp Lys Ser Ile Pro Gly Pro Ser Pro Arg Asn Gln His Arg
 35 40 45

15 (2) INFORMATION FOR SEQ ID NO: 317:

20 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 41 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

25 Met Leu Ile Pro Leu Gln Cys Leu Phe Ser Ser Asp Arg Met Leu Thr
 1 5 10 15

Phe Leu Thr Pro Trp Gln Lys Gly Glu Lys Cys Val Leu Gly Trp Val
 20 25 30

Thr Lys Phe Leu Ser Glu Ile Ser Xaa
 35 40

35 (2) INFORMATION FOR SEQ ID NO: 318:

40 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 76 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

45 Met Thr Phe Ser Ser Leu Lys Leu Phe Val Leu Thr Cys Ile Ile Lys
 1 5 10 15

Gly Leu Glu Arg Phe Ile Ile Leu Arg Glu Val Cys Asn Gln Glu Ile
 20 25 30

Gln Arg Ser Leu Ser Ser Asn Leu Val His Val Leu Leu Gln Pro Ala
 35 40 45

50 Thr Phe Lys Asp Val Leu Val Thr Glu Ile Ile Cys Leu Cys Met Cys
 55 60 65

Leu Tyr Ser Ile Lys Tyr Met Pro Pro Gln Lys Lys
 70 75

60

(2) INFORMATION FOR SEQ ID NO: 339:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

10 Lys Val Tyr Ile Phe Leu Ile Phe Met Val Leu Ile Leu Pro Ser Leu 15

Gly Leu Thr Arg Tyr Met Pro Xaa Ser Xaa Leu Asn Ser Glu 30

15

(2) INFORMATION FOR SEQ ID NO: 340:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

20 Met Ala Lys Ile Ser Pro Phe Glu Val Val Lys Arg Thr Ser Val Pro 15

30 Val Leu Val Gly Leu Val Ile Val Ile Val Ala Thr Glu Leu Met Val 30

Pro Gly Thr Ala Ala Val Thr Gly Lys 40

35

(2) INFORMATION FOR SEQ ID NO: 341:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

40 Met Arg Leu Phe Ile Gly Phe Leu Leu Phe Ser Phe Gly Leu 15

Leu Arg Gln Pro Ser Leu Ser Ala Glu His 25

50

(2) INFORMATION FOR SEQ ID NO: 342:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

60

Met Val Phe Ser Val Ser Ala Leu Ala Leu Leu Met Leu Leu 15

5 Arg Ser Ser Asp Leu Ala Lys Lys Thr Glu 25

(2) INFORMATION FOR SEQ ID NO: 343:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 157 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

10 Met Ser Leu Glu Phe Tyr Gln Lys Lys Ser Arg Trp Pro Phe Ser 15

20 Asp Glu Cys Ile Pro Trp Glu Val Trp Thr Val Lys Val His Val Val 30

Ala Leu Ala Thr Glu Gln Arg Gln Ile Cys Arg Glu Lys Val Gly 45

25 Glu Lys Leu Cys Glu Lys Ile Ile Asn Ile Val Glu Val Met Asn Arg 60

30 His Glu Tyr Leu Pro Lys Met Pro Thr Gln Ser Glu Val Asp Asn Val 80

Phe Asp Thr Gly Leu Arg Asp Val Gln Pro Tyr Leu Tyr Lys Ile Ser 95

35 Phe Gln Ile Thr Asp Ala Leu Gly Thr Ser Val Thr Thr Thr Met Arg 110

Arg Leu Ile Lys Asp Thr Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly 125

40 Ser Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu 140

45 Gly Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp Xaa 155

(2) INFORMATION FOR SEQ ID NO: 344:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 520 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

50 Met Phe Leu Leu Pro Leu Pro Ala Ala Gly Arg Val Val Val Arg Arg 15

60 Leu Ala Val Arg Arg Phe Gly Ser Arg Ser Leu Ser Thr Ala Asp Met

20 25 30
Thr Lys Gly Leu Val Leu Gly Ile Tyr Ser Lys Glu Lys Glu Asp Asp
35 45
5 Val Pro Gln Phe Thr Ser Ala Gly Glu Asn Phe Asp Lys Leu Ala
50 60
10 Gly Lys Leu Arg Glu Thr Leu Asn Ile Ser Gly Pro Pro Leu Lys Ala
65 70 75 80
Gly Lys Thr Arg Thr Phe Tyr Gly Leu His Gln Asp Phe Pro Ser Val
85 90 95
15 Val Leu Val Gly Leu Gly Lys Lys Ala Ala Gly Ile Asp Glu Gln Glu
100 105 110
Asn Trp His Glu Gly Lys Glu Asn Ile Arg Ala Val Ala Gly
115 120 125
20 Cys Arg Gln Ile Gln Asp Leu Glu Leu Ser Ser Val Glu Val Asp Pro
130 135 140
25 Cys Gly Asp Ala Gln Ala Ala Glu Gly Ala Val Leu Gly Leu Tyr
145 150 155 160
Glu Tyr Asp Asp Leu Lys Gln Lys Lys Lys Met Ala Val Ser Ala Lys
165 170 175
30 Leu Tyr Gly Ser Gly Asp Gln Glu Ala Trp Gln Lys Gly Val Leu Phe
180 185 190
35 Ala Ser Gly Gln Asn Leu Ala Arg Gln Leu Met Glu Thr Pro Ala Asn
195 200 205
Glu Met Thr Pro Thr Arg Phe Ala Glu Ile Ile Glu Lys Asn Leu Lys
210 215 220
40 Ser Ala Ser Ser Lys Thr Glu Val His Ile Arg Pro Lys Ser Trp Ile
225 230 235 240
Glu Glu Gln Ala Met Gly Ser Phe Leu Ser Val Ala Lys Gly Ser Asp
245 250 255
45 Glu Pro Pro Val Phe Leu Glu Ile His Tyr Lys Gly Ser Pro Asn Ala
260 265 270
Asn Glu Pro Pro Leu Val Phe Val Gly Lys Gly Ile Thr Phe Asp Ser
275 280 285
50 Gly Gly Ile Ser Ile Lys Ala Ser Ala Asn Met Asp Leu Met Arg Ala
290 295 300
55 Asp Met Gly Gly Ala Thr Ile Cys Ser Ala Ile Val Ser Ala Ala
305 310 315 320
Lys Leu Asn Leu Pro Ile Asn Ile Ile Gly Leu Ala Pro Leu Cys Glu
325 330 335
60 Asn Met Pro Ser Gly Lys Ala Asn Lys Pro Gly Asp Val Val Arg Ala

340 345 350
Lys Asn Gly Lys Thr Ile Gln Val Asp Asn Thr Asp Ala Glu Gly Arg
355 360 365
5 Leu Ile Leu Ala Asp Ala Leu Cys Tyr Ala His Thr Phe Asn Pro Lys
370 375 380
Xaa Ile Leu Asn Ala Ala Thr Leu Thr Gly Ala Met Asp Val Ala Leu
385 390 395 400
Gly Ser Gly Ala Thr Gly Val Phe Thr Asn Ser Ser Trp Leu Trp Asn
405 410 415
15 Lys Leu Phe Glu Ala Ser Ile Glu Thr Gly Asp Arg Val Trp Arg Met
420 425 430
Pro Leu Phe Glu His Tyr Thr Arg Gln Val Val Asp Cys Gln Leu Ala
435 440 445
20 Asp Val Asn Asn Ile Gly Lys Tyr Arg Ser Ala Gly Ala Cys Thr Ala
450 455 460
Ala Ala Phe Leu Lys Glu Phe Val Thr His Pro Lys Trp Ala His Leu
465 470 475 480
Asp Ile Ala Gly Val Met Thr Asn Lys Asp Glu Val Pro Tyr Leu Arg
485 490 495
30 Lys Gly Met Thr Gly Arg Pro Thr Arg Thr Leu Ile Glu Phe Leu Leu
500 505 510
35 Arg Phe Ser Gln Asp Asn Ala Xaa
515 520
(2) INFORMATION FOR SEQ ID NO: 345:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:
45 Thr Ile Leu Phe Leu Phe Leu Gln Leu Ser Ala Leu Arg Leu Ile Val
1 5 10 15
Gly Lys Asp Ser Ile Asp Ile Asp Ile Ser Ser Arg Arg Arg Glu Asp
20 25 30
Gln Ser Leu Arg Leu Asn Ala
35
55 (2) INFORMATION FOR SEQ ID NO: 346:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 234 amino acids
60

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

5 Met Thr Ser Glu Leu Asp Ile Phe Val Gly Asn Thr Thr Leu Ile Asp
1 5 10 15
Glu Asp Val Tyr Arg Leu Trp Leu Asp Gly Tyr Ser Val Thr Asp Ala
20 25 30
Val Ala Leu Arg Val Arg Ser Gly Ile Leu Glu Gln Thr Gly Ala Thr
35 40 45
Ala Ala Val Leu Gln Ser Asp Thr Met Asp His Tyr Arg Thr Phe His
50 55 60
Met Leu Glu Arg Leu His Ala Pro Pro Lys Leu Leu His Gln Leu
65 70 75 80
Ile Phe Gln Ile Pro Pro Ser Arg Gln Ala Leu Leu Ile Glu Arg Tyr
85 90 95
Tyr Ala Phe Asp Glu Ala Phe Val Arg Glu Val Leu Gly Lys Lys Leu
100 105 110
Ser Lys Gly Thr Lys Lys Asp Leu Asp Ile Ser Thr Lys Thr Gly
115 120 125
Ile Thr Leu Lys Ser Cys Arg Arg Gln Phe Asp Asn Phe Lys Arg Val
130 135 140
Phe Lys Val Val Glu Glu Met Arg Gly Ser Leu Val Asp Asn Ile Gln
145 150 155 160
Gln His Phe Leu Leu Ser Asp Arg Leu Ala Arg Asp Tyr Ala Ala Ile
165 170 175
Val Phe Phe Ala Asn Asn Arg Phe Glu Thr Gly Lys Lys Lys Leu Gln
180 185 190
Tyr Leu Ser Phe Gly Asp Phe Ala Phe Cys Ala Glu Leu Met Ile Gln
195 200 205
Asn Trp Thr Leu Gly Pro Val Asp Ser Gln Met Asp Asp Met Asp Met
210 215 220
Asp Leu Asp Arg Asn Phe Ser Arg Thr Xaa
225 230

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 169 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

60 Met Ala Ala Ala Val Ala Gly Met Leu Arg Gly Gly Leu Leu Pro Gln

(2) INFORMATION FOR SEQ ID NO: 347:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 169 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

1 5 10 15
Ala Gly Arg Leu Pro Thr Thr Leu Gln Thr Val Arg Tyr Gly Ser Lys Ala
20 25 30
5 Val Thr Arg His Arg Arg Val Met His Phe Gln Arg Gln Lys Leu Met
35 40 45
Ala Val Thr Glu Tyr Ile Pro Pro Lys Pro Ala Ile His Pro Ser Cys
50 55 60
Leu Pro Ser Pro Ser Pro Gln Glu Ile Gly Leu Ile Arg
65 70 75 80
Leu Leu Arg Arg Glu Ile Ala Ala Val Phe Gln Asp Asn Arg Met Ile
85 90 95
Ala Val Cys Gln Asn Val Ala Leu Ser Ala Glu Asp Lys Leu Leu Ile
100 105 110
Ala Thr Pro Ala Ala Glu Thr Gln Asp Pro Asp Glu Gly Leu Pro Gln
115 120 125
20 Pro Gly Pro Glu Ser Pro Ser Trp Arg Ile Pro Ser Thr Lys Ile Cys
130 135 140
Cys Pro Phe Leu Trp Gly Thr Thr Cys Cys Trp Ser Val Lys Ser Pro
145 150 155 160
30 Arg Ser Arg Arg Trp Tyr Gly Ser Xaa
165

(2) INFORMATION FOR SEQ ID NO: 348:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

Met Lys Arg Ser Phe Leu Leu Pro Leu Leu Val Gly Phe Leu Asp
1 5 10 15
Thr Ala His Leu Ile Leu Leu Glu Thr Leu Ser Val Cys Leu Trp Leu
20 25 30
Pro Ser Leu Ile Asp Ser Arg Cys Val Met Ser
35 40

(2) INFORMATION FOR SEQ ID NO: 349:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

Met Lys Glu Gly Pro Pro Cys Lys Arg His His Tyr Tyr Gln Asn Cys
1 5 10 15

5 Gly Ala Lys Leu Leu Val Ser Leu Phe Gly Glu Thr Asn Gln Ile His
20 25 30

Leu Leu Glu Thr Gln Val Gly Thr Glu Lys Gly Gly Arg Ile Trp
35 40 45

10 Glu Glu Lys Trp Arg Ile Ser Thr Val Leu Phe Ile Ser Val Asn
50 55 60

15 Ser Tyr Val Glu Gly Ser Val Leu Glu Ile Lys Leu Phe Tyr
65 70 75

(2) INFORMATION FOR SEQ ID NO: 350:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

Met Ser Glu Ile Leu Ser Leu Leu Phe Cys Leu Leu Gly Pro Ala Leu
1 5 10 15

30 Asp Glu Arg Arg Glu Lys Asp
20

35 (2) INFORMATION FOR SEQ ID NO: 351:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

Met Ser Ser Ala Gly Thr Ala Thr Pro Leu Glu Met Asp His Lys Leu
1 5 10 15

40 Thr Ser Gln Pro Gly Arg Pro Ser Phe Tyr Cys Asn Ser Arg His Ser
20 25 30

45 Ile Val Gly Ser Ser His Gln Leu Glu Gly Phe Trp Phe Ser His Leu Glu
35 40 45

50 Ser Ser Gly Leu Lys Val Phe Gln Val Ser Leu Pro Cys Glu Cys Val
50 55 60

55 Asn Leu Pro Thr Arg Ile Ala Ser Val Leu Ser Leu Met Ser Leu
65 70 75 80

60 Leu Val Val Gly Gln Ala Pro Ala Trp Glu Gly Ser Leu Leu Arg Gly
85 90 95

Arg Pro Ala Gly Gly Ala His Leu Cys Ala Met Xaa Val Ile Glu Gly
100 105 110

5 Leu Val Val Asp Val Gly Glu Arg Ile Leu His Gly Gln Arg Glu Val
115 120 125

Gly Gln Val Ser Gln Val Leu Pro Ala Leu Ser Leu Gly Leu Val Phe
130 135 140

10 Leu Cys Gln Gly Thr Val Glu Lys Val Ser Gly Ala Ala His Cys Ser
145 150 155 160

Ser Leu Leu Cys Cys Leu Pro Trp Gln Cys Ser Gly Gly Phe Pro
165 170 175

15 Thr Xaa Arg Cys Ser Arg Pro Tyr Phe Ser Ser His Lys Gly Val Ala
180 185 190

20 Ala Thr Leu Ala Leu Thr Cys His Cys Asp Lys Val His Val Ala Gly
195 200 205

Leu Gly Lys Asp Trp Ala Ile Glu Gln Arg Arg Thr Cys Glu Ser
210 215 220

25 Asp Xaa Glu Xaa Xaa Pro Phe Thr Leu Ala Gly Leu Val Leu Val Leu
225 230 235 240

Arg Phe Cys Gln Val Val Leu Val Trp Ile Pro Gln Leu Gly Asp Lys
245 250 255

30 His Trp Arg Gly Met Thr Arg Leu Gly Arg Val Ser Leu Thr Ser Ser
260 265 270

35 Ile Xaa

(2) INFORMATION FOR SEQ ID NO: 352:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

Met Ile Phe Thr Ser Val Thr Lys Gly Ile Leu Leu Ile Ala Leu Trp
1 5 10 15

50 Val Pro Leu Phe His Phe Met Leu Ile Asp Ser Ile Leu Gly Pro Ser
20 25 30

55 Arg Leu Leu Thr Asp Gly Val Pro Phe Asn Pro Trp His Val Xaa
35 40 45

(2) INFORMATION FOR SEQ ID NO: 353:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

5 Met Lys Thr
1

10 (2) INFORMATION FOR SEQ ID NO: 354:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

20 Met Ser Ile Ser Gly Thr Asp Gly Leu Ile Leu Leu Leu Val Gly Leu
1 5 10 15

Glu Ala Xaa Val Arg Ser Ser Lys Lys Trp Ile Pro Lys Ala Leu Xaa
20 25 30

25 Val Thr Gln Ala Lys Trp Asn Ser Trp Pro Ser Arg Asn Ala Gly
35 40 45

30 Phe Ala Leu His
50

(2) INFORMATION FOR SEQ ID NO: 355:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 132 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

40 Met Glu His Cys Leu Tyr His Ser Val His Gly Ile Asn Pro Tyr Ile
1 5 10 15

45 His Lys Asn Thr His Pro Ser Ile Asn Ile Tyr Met Val Trp Asp Glu
20 25 30

Gln Val Asn Ser Phe Glu Arg Glu Phe Val Pro Phe Phe Leu Ile
35 40 45

50 Ile Leu Leu Asn Cys Cys Gln Leu Ser Asn Lys Gln Thr Glu Lys Leu
50 55 60

Phe Gly Lys Thr Leu His Thr Pro Phe Leu Ser Ser Ala Leu Lys Tyr
65 70 75 80

55 Arg Leu Asn Thr His Ile Leu Pro Val Phe Ser Tyr Ser Asp Ser Ile
85 90 95

60 Leu Thr Cys His Leu Ile Leu Ala Ser Tyr Phe Ser His Val Tyr Leu
100 105 110

Pro Val Thr Cys Ile Cys Tyr Leu Asn Arg Lys Lys Asn Ile Gln Lys
115 120 125

5 Lys Lys Asn Xaa
130

10 (2) INFORMATION FOR SEQ ID NO: 356:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 204 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

20 Met Gly Ser Arg Asp His Leu Phe Lys Val Leu Val Val Gly Asp Ala
1 5 10 15

Ala Val Gly Lys Thr Ser Leu Val Gln Asp Tyr Ser Gln Asn Ser Phe
20 25 30

25 Ser Lys His Tyr Lys Ser Thr Val Gly Val Asp Phe Ala Leu Lys Val
35 40 45

Leu Gln Trp Ser Asp Tyr Glu Ile Val Arg Leu Gln Leu Trp Asp Ile
50 55 60

30 Ala Gly Gln Glu Arg Phe Thr Ser Met Thr Arg Leu Tyr Tyr Arg Asp
65 70 75 80

35 Ala Ser Ala Cys Val Ile Met Phe Asp Val Thr Asn Ala Thr Thr Phe
85 90 95

Ser Asn Ser Gln Arg Trp Lys Gln Asp Leu Asp Ser Lys Leu Thr Leu
100 105 110

40 Pro Asn Gly Glu Pro Val Pro Cys Leu Leu Ala Asn Lys Cys Asp
115 120 125

Leu Ser Pro Trp Ala Val Ser Arg Asp Gln Ile Asp Arg Phe Ser Lys
130 135 140

45 Glu Asn Gly Phe Thr Gly Trp Thr Glu Thr Ser Val Lys Glu Asn Lys
145 150 155 160

50 Asn Ile Asn Glu Ala Met Arg Val Leu Ile Glu Lys Met Met Arg Asn
165 170 175

Ser Thr Glu Asp Ile Met Ser Leu Ser Thr Gln Gly Asp Tyr Ile Asn
180 185 190

55 Leu Gln Thr Lys Ser Ser Trp Ser Cys Xaa
195 200

(2) INFORMATION FOR SEQ ID NO: 357:

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

Met Ile Ser Leu Ile Phe Gln Glu Glu Glu Lys Leu Val Glu Lys
1 5 10 15

10 Phe Phe Phe Leu Phe Phe Leu Lys Lys Gly Ser Gln Gly Ser
20 25 30

Asn Leu Lys Ile Val Pro Arg His Met Arg Val Val Leu Arg Gly
35 40 45

(2) INFORMATION FOR SEQ ID NO: 358:

(A) LENGTH: 73 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

25 Met Thr Tyr Val Thr Cys Leu His Val Cys Leu Leu Val Glu Phe Leu
1 5 10 15

30 Asn Ser Gln Leu Thr Asn His Arg Lys Tyr Tyr Phe Leu Ser Tyr Gly
20 25 30

Phe Trp Phe Thr Gly Leu Arg Gly Phe Ser Glu Tyr Trp Trp Pro Gln
35 40 45

35 Gln His Thr Ser Phe His Pro Asn Arg Asn Glu Ile Asn Phe Val Ser
50 55 60

Thr Asp Asn Arg Ile Trp Val Thr Xaa
65 70

(2) INFORMATION FOR SEQ ID NO: 359:

(A) LENGTH: 102 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

45 Met Ser Asp Gln Glu Ala Lys Pro Ser Thr Glu Asp Leu Gly Asp Lys
1 5 10 15

50 Lys Glu Gly Glu Tyr Ile Lys Leu Lys Val Ile Gly Gln Asp Ser Ser
20 25 30

55 Glu Ile His Phe Lys Val Lys Met Thr Thr His Leu Lys Lys Leu Lys
35 40 45

60 Glu Ser Tyr Cys Gln Arg Gln Gly Val Pro Met Asn Ser Leu Arg Phe

50 55 60

Leu Phe Glu Gly Gln Arg Ile Ala Asp Asn His Thr Pro Lys Glu Leu
65 70 75 80

5 Gly Met Glu Glu Asp Val Ile Glu Val Tyr Gln Glu Gln Thr Gly
85 90 95

10 Gly His Ser Thr Val Xaa
100

(2) INFORMATION FOR SEQ ID NO: 360:

(A) LENGTH: 48 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

20 Met Gly Phe Pro Gln Trp His Leu Gly Asn His Ala Val Glu Pro Val
1 5 10 15

25 Thr Ser Ile Leu Leu Leu Phe Leu Leu Met Met Leu Gly Val Arg Gly
20 25 30

30 Leu Leu Leu Val Gly Leu Val Tyr Leu Val Ser His Leu Ser Gln Arg
35 40 45

(2) INFORMATION FOR SEQ ID NO: 361:

(A) LENGTH: 179 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

45 Met Ser Ala Glu Val Lys Val Thr Gly Gln Asn Gln Glu Gln Phe Leu
1 5 10 15

Leu Leu Ala Lys Ser Ala Lys Gly Ala Ala Leu Ala Thr Leu Ile His
20 25 30

50 Gln Val Leu Glu Ala Pro Gly Val Tyr Val Phe Gly Glu Leu Leu Asp
35 40 45

Met Pro Asn Val Arg Glu Leu Ala Glu Ser Asp Phe Ala Ser Thr Phe
50 55 60

55 Arg Leu Leu Thr Val Phe Ala Tyr Gly Thr Tyr Ala Asp Tyr Leu Ala
65 70 75 80

60 Glu Ala Arg Asn Leu Pro Leu Thr Glu Ala Gln Lys Asn Lys Leu
85 90 95

Arg His Leu Ser Val Val Thr Leu Ala Ala Lys Val Lys Cys Ile Pro
100 105 110

5 Tyr Ala Val Leu Leu Glu Ala Leu Ala Leu Arg Asn Val Arg Gln Leu
115 120 125

Glu Asp Leu Val Ile Glu Ala Val Tyr Ala Asp Val Leu Arg Gly Ser
130 135 140

10 Leu Asp Gln Arg Asn Gln Arg Leu Glu Val Asp Tyr Ser Ile Gly Arg
145 150 155 160

Asp Ile Gln Arg Gln Asp Leu Ser Ala Ile Ala Arg Thr Leu Xaa Lys
165 170 175

Asn His Xaa

(2) INFORMATION FOR SEQ ID NO: 362:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

30 Met Lys Ser Ser Ser Leu Phe Phe Phe Leu Ala His Phe Ile His
1 5 10 15

Ser His Asp Leu Pro Gly Leu Cys Arg
20 25

(2) INFORMATION FOR SEQ ID NO: 363:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

45 Met Lys Phe Ala Ala Ser Gly Xaa Phe Leu His His Met Ala Gly Leu
1 5 10 15

Ser Ser Lys Leu Ser Met Ser Lys Ala Leu Pro Leu Thr Lys Val
20 25 30

Val Gln Asn Asp Ala Tyr Thr Ala Pro Ala Leu Pro Ser Ile Arg
35 40 45

55 Thr Lys Ala Leu Thr Asn Met Ser Arg Thr Leu Val Asn Lys Glu
50 55 60

Pro Pro Lys Glu Leu Pro Ala Ala Glu Pro Val Leu Ser Pro Leu Glu
65 70 75 80

Gly Thr Lys Met Thr Val Asn Asn Leu His Pro Arg Val Thr Glu Glu
85 90 95

5 Asp Ile Val Glu Leu Phe Cys Val Cys Gly Ala Leu Lys Arg Ala Arg
100 105 110

Leu Val His Pro Gly Val Ala Glu Val Val Phe Val Lys Lys Asp Asp
115 120 125

10 Ala Ile Thr Ala Tyr Lys Lys Tyr Asn Asn Arg Cys Leu Asp Gly Gln
130 135 140

Pro Met Lys Cys Asn Leu His Met Asn Gly Asn Val Ile Thr Ser Asp
145 150 155 160

15 Gln Pro Ile Leu Leu Arg Leu Ser Asp Ser Pro Ser Met Lys Lys Glu
165 170 175

Ser Glu Leu Pro Arg Arg Val Asn Ser Ala Ser Ser Asn Pro Pro
180 185 190

Ala Glu Val Asp Pro Asp Thr Ile Leu Lys Ala Leu Phe Lys Ser Ser
195 200 205

25 Gly Ala Ser Xaa Thr Thr Gln Pro Thr Glu Phe Lys Ile Lys Leu Xaa
210 215 220

(2) INFORMATION FOR SEQ ID NO: 364:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 349 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

40 Met Ser Lys Asn Cys Ile Lys Leu Leu Cys Glu Asp Pro Val Phe Ala
1 5 10 15

Glu Tyr Ile Lys Cys Ile Leu Met Asp Glu Arg Thr Phe Leu Asn Asn
20 25 30

Asn Ile Val Tyr Thr Phe Met Thr Thr Phe Leu Leu Lys Val Gln Ser
35 40 45

50 Gln Val Phe Ser Glu Ala Asn Cys Ala Asn Leu Ile Ser Thr Leu Ile
50 55 60

Thr Asn Leu Ile Ser Gln Tyr Gln Asn Leu Gln Ser Asp Phe Ser Asn
65 70 75 80

Arg Val Glu Ile Ser Lys Ala Ser Ala Ser Leu Asn Gly Asp Leu Arg
85 90 95

60 Ala Leu Ala Leu Leu Ser Val His Thr Pro Lys Gln Leu Asn Pro
100 105 110

Ala Leu Ile Pro Thr Leu Gln Glu Leu Ser Lys Cys Arg Thr Cys
115 120 125

5 Leu Gln Gln Arg Asn Ser Leu Gln Gln Glu Ala Lys Glu Arg Lys
130 135 140

Thr Lys Asp Asp Glu Gly Ala Thr Pro Ile Lys Arg Arg Val Ser
145 150 155 160

10 Ser Asp Glu Glu His Thr Val Asp Ser Cys Ile Ser Asp Met Lys Thr
165 170 175

15 Glu Thr Arg Glu Val Leu Thr Pro Thr Ser Thr Ser Asp Asn Glu Thr
180 185 190

Arg Asp Ser Ser Ile Ile Asp Pro Gly Thr Glu Gln Asp Leu Pro Ser
195 200 205

20 Pro Glu Asn Ser Ser Val Lys Glu Tyr Arg Met Glu Val Pro Ser Ser
210 215 220

Phe Ser Glu Asp Met Ser Asn Ile Arg Ser Gln His Ala Glu Gln
225 230 235 240

25 Ser Asn Asn Gly Arg Tyr Asp Asp Cys Lys Glu Phe Lys Asp Leu His
245 250 255

Cys Ser Lys Asp Ser Thr Leu Ala Glu Glu Ser Glu Phe Pro Ser
260 265 270

30 Thr Ser Ile Ser Ala Val Leu Ser Asp Leu Ala Asp Leu Arg Ser Cys
275 280 285

35 Asp Gly Gln Ala Leu Pro Ser Gln Asp Pro Glu Val Ala Leu Ser Leu
290 295 300

Ser Cys Gly His Ser Arg Gly Leu Phe Ser His Met Gln Gln His Asp
305 310 315 320

40 Ile Leu Asp Thr Leu Cys Arg Thr Ile Glu Ser Thr Ile His Val Val
325 330 335

45 Thr Arg Ile Ser Gly Lys Gly Asn Gln Ala Ala Ser Xaa
340 345

50 (2) INFORMATION FOR SEQ ID NO: 365:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 467 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 365:
Met Leu His Gln Asp His Ile Thr Phe Ala Met Leu Leu Ala Arg Ile
1 5 10 15

60 Lys Leu Lys Gly Thr Val Gly Glu Pro Thr Tyr Asp Ala Glu Phe Gln

20 25 30
His Phe Leu Arg Gly Asn Glu Ile Val Leu Ser Ala Gly Ser Thr Pro
35 40 45

5 Arg Ile Gln Gly Leu Thr Val Glu Gln Ala Glu Ala Val Val Arg Leu
50 55 60

10 Ser Cys Leu Pro Ala Phe Lys Asp Leu Ile Ala Lys Val Gln Ala Asp
65 70 75 80

Glu Gln Phe Gly Ile Thr Leu Asp Ser Ser Pro Glu Gln Thr Val
85 90 95

15 Pro Tyr Leu Thr Ser Glu Glu Thr Pro Ala Thr Pro Ile Gly Gln Ala
100 105 110

Ile His Arg Leu Leu Ile Gln Ala Phe Arg Pro Asp Arg Leu Leu
115 120 125

20 Ala Met Ala His Met Phe Val Ser Thr Asn Leu Gly Glu Ser Phe Met
130 135 140

Ser Ile Met Glu Gln Pro Leu Asp Leu Thr His Ile Val Xaa Thr Glu
145 150 155 160

Val Lys Pro Asn Thr Pro Val Leu Met Cys Ser Val Pro Gly Tyr Asp
165 170 175

30 Ala Ser Gly His Val Glu Asp Leu Ala Glu Gln Asn Thr Gln Ile
180 185 190

Thr Ser Ile Ala Ile Gly Ser Ala Glu Gly Phe Asn Gln Ala Asp Lys
195 200 205

35 Ala Ile Asn Thr Ala Val Lys Ser Gly Arg Trp Val Met Leu Lys Asn
210 215 220

Val His Leu Ala Pro Gly Trp Leu Met Gln Leu Glu Lys Lys Leu His
225 230 235 240

Ser Leu Gln Pro His Ala Cys Phe Arg Leu Phe Leu Thr Met Glu Ile
245 250 255

45 Asn Pro Lys Val Pro Val Asn Leu Leu Arg Ala Gly Arg Ile Phe Val
260 265 270

Phe Glu Pro Pro Gly Xaa Lys Ala Asn Met Leu Arg Thr Phe Ser
275 280 285

50 Ser Ile Pro Val Ser Arg Ile Cys Lys Ser Pro Asn Glu Arg Ala Arg
290 295 300

Leu Tyr Phe Leu Leu Ala Trp Phe His Ala Ile Ile Gln Glu Arg Leu
305 310 315 320

Arg Tyr Ala Pro Leu Gly Trp Ser Lys Lys Tyr Glu Phe Gly Glu Ser
325 330 335

60 Asp Leu Arg Ser Xaa Cys Asp Thr Val Asp Thr Trp Leu Asp Asp Thr

340 345 350
 Ala Lys Gly Arg Gln Asn Ile Ser Pro Asp Lys Ile Pro Trp Ser Ala
 355 360 365
 5 Leu Lys Thr Leu Met Ala Gln Ser Ile Tyr Gly Gly Arg Val Asp Asn
 370 375 380
 10 Glu Phe Asp Gln Arg Leu Leu Asn Thr Phe Leu Glu Arg Leu Phe Thr
 385 390 395
 Thr Arg Ser Phe Asp Ser Glu Phe Lys Leu Ala Cys Lys Val Asp Gly
 400 405 410 415
 15 His Lys Asp Ile Gln Met Pro Asp Gly Met Gln Ala Arg Gly Val Cys
 420 425 430
 Ala Val Gly Gly Val Ala Pro Arg His Pro Asp Ala Leu Leu Ala Gly
 435 440 445
 20 Pro Ala Gln Gln Arg Arg Glu Ser Pro Pro Tyr His Thr Gly Cys Gly
 450 455 460
 25 His Asp Gln
 465
 30 (2) INFORMATION FOR SEQ ID NO: 366:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 152 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 366:
 Met Ala Asp Glu Ala Thr Arg Arg Val Val Ser Glu Ile Pro Val Leu
 1 5 10 15
 Lys Thr Asn Ala Gly Pro Arg Asp Arg Glu Leu Trp Val Gln Arg Leu
 20 25 30
 Lys Glu Glu Tyr Gln Ser Leu Ile Arg Tyr Val Glu Asn Asn Lys Asn
 35 40 45
 Ala Asp Asn Asp Trp Phe Arg Leu Glu Ser Asn Lys Glu Gly Thr Arg
 50 55 60
 Trp Phe Gly Lys Cys Trp Tyr Ile His Asp Leu Lys Tyr Glu Phe
 65 70 75 80
 Asp Ile Glu Phe Asp Ile Pro Ile Thr Tyr Pro Thr Thr Ala Pro Glu
 85 90 95
 55 Ile Ala Val Pro Glu Leu Asp Gly Lys Thr Ala Lys Met Tyr Arg Gly
 100 105 110
 Gly Lys Ile Cys Leu Thr Asp His Phe Lys Pro Leu Trp Gly Gln Glu
 115 120 125
 60

Cys Ala Gln Ile Trp Thr Ser Ser Ser His Gly Ser Gly Ala Gly Ser
 130 135 140
 5 Met Xaa Gly Ser Gly Asn Pro Xaa
 145 150
 10 (2) INFORMATION FOR SEQ ID NO: 367:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 373 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:
 Met Tyr Asp Gly Thr Lys Glu Val Pro Met Asn Pro Val Lys Ile Tyr
 1 5 10 15
 Gln Val Cys Asp Ile Pro Gln Pro Gln Gly Ser Ile Ile Asn Pro Gly
 20 25 30
 Ser Thr Gly Ser Ala Pro Trp Asp Glu Lys Asp Asn Asp Val Asp Glu
 35 40 45
 Glu Asp Glu Glu Asp Glu Leu Asp Gln Ser Gln His His Val Pro Ile
 50 55 60
 30 Gln Asp Thr Phe Pro Phe Leu Asn Ile Asn Gly Ser Pro Met Ala Pro
 65 70 75 80
 Ala Ser Val Gly Asn Cys Ser Val Gly Asn Cys Ser Pro Glu Ala Val
 85 90 95
 35 Trp Pro Lys Thr Glu Pro Leu Glu Met Glu Val Pro Gln Ala Pro Ile
 100 105 110
 Gln Pro Phe Tyr Ser Ser Pro Glu Leu Trp Ile Ser Ser Leu Pro Met
 115 120 125
 40 Thr Asp Leu Asp Ile Lys Phe Gln Tyr Arg Gly Lys Glu Tyr Gly Gln
 130 135 140 145
 Thr Met Thr Val Ser Asn Pro Gln Gly Cys Arg Leu Phe Thr Gly Asp
 145 150 155 160
 Leu Gly Pro Met Pro Asp Gln Glu Leu Phe Gly Pro Val Xaa Leu
 165 170 175
 50 Glu Gln Val Lys Phe Pro Gly Pro Glu His Ile Thr Asn Glu Lys Gln
 180 185 190
 Lys Leu Phe Thr Ser Lys Leu Leu Asp Val Met Asp Arg Gly Leu Ile
 195 200 205
 55 Leu Glu Val Ser Gly His Ala Ile Tyr Ala Ile Arg Leu Cys Gln Cys
 210 215 220 225
 Lys Val Tyr Trp Ser Gly Pro Cys Ala Pro Ser Leu Val Ala Pro Asn
 230 235 240

Leu Ile Glu Arg Gln Lys Lys Val Lys Leu Phe Cys Leu Glu Thr Phe
245 250 255

5 Leu Ser Asp Leu Ile Ala His Gln Lys Gly Gln Ile Glu Lys Gln Pro
260 265 270

Pro Phe Glu Ile Tyr Leu Cys Phe Gly Glu Glu Trp Pro Asp Gly Lys
275 280 285

10 Pro Leu Glu Arg Lys Leu Ile Leu Val Gln Val Ile Pro Val Val Ala
290 295 300

Arg Met Ile Tyr Glu Met Phe Ser Gly Asp Phe Thr Arg Ser Phe Asp
305 310 315 320

Ser Gly Ser Val Arg Leu Gln Ile Ser Thr Pro Asp Ile Lys Asp Asn
325 330 335

20 Ile Val Ala Gln Leu Lys Gln Leu Tyr Arg Ile Leu Gln Thr Gln Glu
340 345 350

Ser Trp Gln Pro Met Gln Pro Thr Pro Ser Met Gln Leu Pro Pro Ala
355 360 365

25 Leu Pro Pro Gln Xaa
370

30 (2) INFORMATION FOR SEQ ID NO: 368:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser
1 5 10 15

Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe
20 25 30

45 Trp Gly Phe Gly Arg Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr
35 40 45

Pro Pro Asp Gln Glu Pro Gln Lys Glu Thr Glu Pro Ala Thr Ser Ser
50 55 60

50 His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln
65 70 75 80

Pro Asn Xaa

(2) INFORMATION FOR SEQ ID NO: 369:

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala
1 5 10 15

10 Tyr Trp Thr Met Xaa
20

15 (2) INFORMATION FOR SEQ ID NO: 370:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 227 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

Met Gly Ala Ser Ala Arg Leu Leu Arg Ala Val Ile Met Gly Ala Pro
1 5 10 15

25 Gly Ser Gly Lys Gly Thr Val Ser Ser Arg Ile Thr Thr His Phe Glu
20 25 30

30 Leu Lys His Leu Ser Ser Gly Asp Leu Leu Arg Asp Asn Met Leu Arg
35 40 45

Gly Thr Glu Ile Gly Val Leu Ala Lys Ala Phe Ile Asp Gln Gly Lys
50 55 60

35 Leu Ile Pro Asp Asp Val Met Thr Arg Leu Ala Leu His Glu Leu Lys
65 70 75 80

Asn Leu Thr Gln Tyr Ser Trp Leu Leu Asp Gly Phe Pro Arg Thr Leu
85 90 95

40 Pro Gln Ala Glu Ala Leu Asp Arg Ala Tyr Gln Ile Asp Thr Val Ile
100 105 110

45 Asn Leu Asn Val Pro Phe Glu Val Ile Lys Gln Arg Leu Thr Ala Arg
115 120 125

Trp Ile His Pro Ala Ser Gly Arg Val Tyr Asn Ile Glu Phe Asn Pro
130 135 140

50 Pro Lys Thr Val Gly Ile Asp Asp Leu Thr Gly Glu Pro Leu Ile Gln
145 150 155 160

Arg Glu Asp Asp Lys Pro Glu Thr Val Ile Lys Arg Leu Lys Ala Tyr
165 170 175

Glu Asp Gln Thr Lys Pro Val Leu Glu Tyr Tyr Gln Lys Lys Gly Val
180 185 190

60 Leu Glu Thr Phe Ser Gly Thr Glu Thr Asn Lys Ile Trp Pro Tyr Val
195 200 205

Tyr Ala Phe Leu Gln Thr Lys Val Pro Gln Arg Ser Gln Lys Ala Ser
210 215 220

5 Val Thr Pro
225

10 (2) INFORMATION FOR SEQ ID NO: 371:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

Met Phe Leu Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Gln
1 5 10 15

Leu Lys Ile Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu
20 25 30

Lys Ile Ile Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Ser
35 40 45

Trp Ala Ile Lys Ala Gln Leu Lys Ile Glu Asn Lys Asp Leu Asp Asn
50 55 60

Lys Thr Ala Lys Gly Gly Gln Glu Ala Leu Thr Cys Thr Xaa
65 70 75

35 (2) INFORMATION FOR SEQ ID NO: 372:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

Phe Arg Ala Val Phe Pro Cys Cys Pro Phe Leu Thr Leu Met Leu Pro
1 5 10 15

Leu Leu Glu Cys Leu Val Gly Met Ile Met Cys Tyr Leu Gly Ile Ser
20 25 30

Phe Thr Asp Thr Arg Lys Thr Ala Gly Leu Lys Lys Lys Lys Lys
35 40 45

Lys Xaa Xaa
50

(2) INFORMATION FOR SEQ ID NO: 373:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

5 Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser
1 5 10 15

Phe Ile Val Thr Pro Thr Ser Asn Thr Ala Ser Tyr Leu Trp
20 25 30

Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Ser Trp Ala Cys
35 40 45

Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu Xaa
50 55 60

20 (2) INFORMATION FOR SEQ ID NO: 374:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe
1 5 10 15

Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg
20 25 30

Ile Leu Phe Phe Ile Val Phe Xaa
35 40

(2) INFORMATION FOR SEQ ID NO: 375:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

Met Cys Ser Gly Gln Ser Gln Val Trp Lys Met Ala Leu Gln Ala Leu
1 5 10 15

Asp Ser Glu Thr Val Val Ile Leu Pro Asp Met His Leu Ile Leu Ser
20 25 30

Leu Arg Leu Ile His Asn Ala Arg Pro Cys Leu Xaa
35 40

(2) INFORMATION FOR SEQ ID NO: 376:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 203 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

5 Met Leu Ile Ser Glu Glu Ile Pro Phe Lys Asp Asp Pro Arg Asp 15
1 5 10
Glu Thr Tyr Lys Pro His Leu Glu Arg Glu Thr Pro Lys Pro Arg Arg 30
20 25
Lys Ser Gly Lys Val Lys Glu Glu Lys Glu Lys Lys Glu Ile Lys Val 45
35 40
Glu Val Glu Val Glu Val Lys Glu Glu Glu Aun Glu Ile Arg Glu Asp 60
50 55
Glu Glu Pro Pro Arg Lys Arg Gly Arg Arg Arg Lys Asp Lys Ser 75 80
65 70
Pro Arg Leu Pro Lys Arg Arg Lys Lys Pro Pro Ile Gln Tyr Val Arg 95
85 90
Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His Pro Arg Tyr Leu 110
100 105
Gln His Ile Lys Tyr Gln His Leu Leu Lys Lys Tyr Val Cys 125
115 120
Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln Lys Gln Leu Leu 140
130 135
Arg His Ala Lys His Thr Asp Gln Arg Asp Tyr Ile Cys Glu Tyr 155
145 150
Cys Ala Arg Ala Phe Lys Ser Ser His Aun Leu Ala Val His Arg Met 175
165 170
Ile His Thr Gly Glu Lys His Tyr Aun Val Arg Ser Val Asp Leu Leu 190
180 185
Val Asp Lys Arg His Leu Leu Ile Gly Thr Xaa 200
195 200

(2) INFORMATION FOR SEQ ID NO: 377:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

50 Met Leu Pro Arg Arg Thr Phe Tyr Phe Thr Phe Ile Phe Ile Phe Phe 15
1 5 10
Leu Ala Ser Phe Thr Gly Phe Thr Leu Arg Ala Ser Phe 25
20 25

(2) INFORMATION FOR SEQ ID NO: 378:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 136 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

10 Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Leu Met 15
1 5 10
Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val Leu 30
20 25
Tyr Leu His Aun His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu Trp 45
35 40
Asp Ser Phe Aun Glu Val Thr Aun Gln Thr Leu Asp Val Lys Arg Met 60
50 55
Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val Gln 80
65 70 75
Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu Aun 95
85 90
Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met Pro Ser Phe 110
100 105
Phe Ser Cys His Leu Phe Cys Thr Leu Arg Trp Lys Tyr Phe Glu Val 125
115 120
Phe Tyr Aun His Lys Phe Leu Xaa 135
130 135

(2) INFORMATION FOR SEQ ID NO: 379:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

45 Met Ala Trp Arg Arg Arg Glu Pro Ala Ser Gly Leu Ala Ala Cys Trp 15
1 5 10
Leu Trp Arg Cys Ser Pro Trp Pro Cys Ala Cys Pro Gly Pro Gly Ala 30
20 25
Gly Leu Ser Ser Gly Ser Arg Pro Trp 40
35 40

(2) INFORMATION FOR SEQ ID NO: 380:

(1) SEQUENCE CHARACTERISTICS:

580

(A) LENGTH: 468 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

5 Met Glu Phe Leu Lys Val Ala Arg Arg Asn Lys Arg Glu Gln Leu Glu 15
 1 5 10 15
 10 Gln Ile Gln Lys Glu Leu Ser Val Leu Glu Glu Asp Ile Lys Arg Val 30
 20 25 30
 Glu Glu Met Ser Gly Leu Tyr Ser Pro Val Ser Glu Asp Ser Thr Val 45
 35 45
 15 Pro Gln Phe Glu Ala Pro Ser Pro Ser His Ser Ser Ile Ile Asp Ser 60
 50 55 60
 Thr Glu Tyr Ser Gln Pro Pro Gly Phe Ser Gly Ser Ser Gln Thr Lys 80
 65 70 75 80
 20 Lys Gln Pro Trp Tyr Asn Ser Thr Leu Ala Ser Arg Arg Lys Arg Leu 95
 85 90 95
 Thr Ala His Phe Glu Asp Leu Glu Gln Cys Tyr Phe Ser Thr Arg Met 110
 100 105 110
 Ser Arg Ile Ser Asp Asp Ser Arg Thr Ala Ser Gln Leu Asp Glu Phe 125
 115 120 125
 30 Gln Glu Cys Leu Ser Lys Phe Thr Arg Tyr Asn Ser Val Arg Pro Leu 140
 130 135 140
 Ala Thr Leu Ser Tyr Ala Ser Asp Leu Tyr Asn Gly Ser Ser Ile Val 160
 145 150 155 160
 35 Ser Ser Ile Glu Phe Asp Arg Asp Cys Asp Tyr Phe Ala Ile Ala Gly 175
 165 170 175
 Val Thr Lys Lys Ile Lys Val Tyr Glu Tyr Asp Thr Val Ile Gln Asp 190
 180 185 190
 Ala Val Asp Ile His Tyr Pro Glu Asn Glu Met Thr Cys Asn Ser Lys 205
 195 200 205
 45 Ile Ser Cys Ile Ser Trp Ser Ser Tyr His Lys Asn Leu Leu Ala Ser 220
 210 215 220
 Ser Asp Tyr Glu Gly Thr Val Ile Leu Trp Asp Gly Phe Thr Gly Gln 240
 225 230 235 240
 50 Arg Ser Lys Val Tyr Gln Glu His Glu Lys Arg Cys Trp Ser Val Asp 255
 245 250 255
 Phe Asn Leu Met Asp Pro Lys Leu Leu Ala Ser Gly Ser Asp Asp Ala 270
 260 265 270
 Lys Val Lys Leu Trp Ser Thr Asn Leu Asp Asn Ser Val Ala Ser Ile 285
 275 280 285
 60 Glu Ala Lys Ala Asn Val Cys Cys Val Lys Phe Ser Pro Ser Ser Arg

581

290 295 300

5 Tyr His Leu Ala Phe Gly Cys Ala Asp His Cys Val His Tyr Tyr Asp 320
 305 310 315 320
 Leu Arg Asn Thr Lys Gln Pro Ile Met Val Phe Lys Gly His Arg Lys 335
 325 330 335
 10 Ala Val Ser Tyr Ala Lys Phe Val Ser Gly Glu Glu Ile Val Ser Ala 350
 340 345 350
 Ser Thr Asp Ser Gln Leu Lys Leu Trp Asn Val Gly Lys Pro Tyr Cys 365
 355 360 365
 15 Leu Arg Ser Phe Lys Gly His Ile Asn Glu Lys Asn Phe Val Gly Leu 380
 370 375 380
 Ala Ser Asn Gly Asp Tyr Ile Ala Cys Gly Ser Glu Asn Asn Ser Leu 400
 385 390 395 400
 20 Tyr Leu Tyr Tyr Lys Gly Leu Ser Lys Thr Leu Leu Thr Phe Lys Phe 415
 405 410 415
 Asp Thr Val Lys Ser Val Leu Asp Lys Asp Arg Lys Glu Asp Asp Thr 430
 420 425 430
 Asn Glu Phe Val Ser Ala Val Cys Trp Arg Ala Leu Pro Asp Gly Glu 445
 435 440 445
 30 Ser Asn Val Leu Ile Ala Ala Asn Ser Gln Gly Thr Ile Lys Val Leu 460
 450 455 460
 Glu Leu Val Xaa 465
 465
 35 (2) INFORMATION FOR SEQ ID NO: 381:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:
 45 Met Arg Lys Glu Asp Gly Phe Thr Phe Phe Phe Leu Phe Phe Phe 15
 1 5 10 15
 Val Val Gly Ser Lys Phe Val Asn Gly Asn Lys Leu Val 25
 20 25
 55 (2) INFORMATION FOR SEQ ID NO: 382:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 382:

Met Pro Leu Ala Pro Tyr Cys Asp Leu Leu Val Ala Leu Ser Phe Ala
1 10 15

5 Leu Val Leu Glu Ser Pro Val Asp Ser Ser Asp Phe Thr
20 25

10 (2) INFORMATION FOR SEQ ID NO: 383:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 138 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

Met Asn Ser Leu Val Ser Trp Gln Leu Leu Phe Leu Cys Ala Thr
1 5 10 15

20 His Phe Gly Glu Pro Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg
25 30

25 Pro Thr Gly Gln Gln Leu Glu Ser Leu Gly Leu Ala Pro Gly Glu
35 40 45

Gln Ser Leu Pro Cys Thr Glu Arg Lys Pro Ala Ala Thr Ala Arg Leu
50 55 60

30 Ser Arg Arg Gly Thr Ser Leu Ser Pro Pro Glu Ser Ser Gly Ser
65 70 75 80

35 Pro Gln Gln Pro Gly Leu Ser Ala Pro His Ser Arg Gln Ile Pro Ala
85 90 95

Pro Gln Gly Ala Val Leu Val Gln Arg Glu Lys Asp Leu Pro Asn Tyr
100 105 110

40 Asn Trp Asn Ser Phe Gly Leu Arg Phe Gly Lys Arg Glu Ala Ala Pro
115 120 125

Gly Asn His Gly Arg Ser Ala Gly Arg Gly
130 135

(2) INFORMATION FOR SEQ ID NO: 384:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

Met Ser Cys Phe Ile Asp Ser Xaa Asp Ser Lys Ile Leu His Leu Leu
1 5 10 15

Val Val Ser Phe Ile Cys Xaa Leu Phe Leu Leu Ile Leu Thr His Gly
20 25 30

Ile Leu Ile Leu Arg Xaa Phe Ser Val Xaa Xaa His Ser Leu Lys
35 40 45

5 Asn Asn Leu Glu Glu Tyr Leu Ile Leu Met Asn Lys Ala Leu Leu Thr
50 55 60

Arg Glu Asp Phe Phe Val Leu Pro Xaa Ala
65 70

10

(2) INFORMATION FOR SEQ ID NO: 385:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 521 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 385:

20 Met Ser Ala Gly Glu Val Glu Arg Leu Val Ser Glu Leu Ser Gly Gly
1 5 10 15

25 Thr Gly Gly Asp Glu Glu Glu Trp Leu Tyr Gly Asp Glu Asn Glu
30 35

Val Glu Arg Pro Glu Glu Asn Ala Ser Ala Asn Pro Pro Ser Gly
40 45

30 Ile Glu Asp Glu Thr Ala Glu Asn Gly Val Pro Lys Pro Lys Val Thr
50 55 60

Glu Thr Glu Asp Asp Ser Asp Ser Asp Asp Glu Asp Glu Asp
65 70 75 80

35 Val His Val Thr Ile Gly Asp Ile Lys Thr Gly Ala Pro Gln Tyr Gly
85 90 95

Ser Tyr Gly Thr Ala Pro Val Asn Leu Asn Ile Lys Thr Gly Gly Arg
100 105 110

40 Val Tyr Gly Thr Thr Gly Thr Lys Val Lys Gly Val Asp Leu Asp Ala
115 120 125

45 Pro Gly Ser Ile Asn Gly Val Pro Leu Leu Glu Val Asp Leu Asp Ser
130 135 140

Phe Glu Asp Lys Pro Trp Arg Lys Pro Gly Ala Asp Leu Ser Asp Tyr
145 150 155 160

50 Phe Asn Tyr Gly Phe Asn Glu Asp Thr Trp Lys Ala Tyr Cys Glu Lys
165 170 175

Gln Lys Arg Ile Arg Met Gly Leu Glu Val Ile Pro Val Thr Ser Thr
180 185 190

55 Thr Asn Lys Ile Thr Val Gln Gln Gly Arg Thr Gly Asn Ser Glu Lys
195 200 205

60 Glu Thr Ala Leu Pro Ser Thr Lys Ala Glu Phe Thr Ser Pro Pro Ser
210 215 220

Leu Phe Lys Thr Gly Leu Pro Pro Ser Arg Arg Leu Pro Gly Ala Ile
225 230 235 240
5 Asp Val Ile Gly Gln Thr Ile Thr Ile Ser Arg Val Glu Gly Arg
245 250 255
Arg Ala Asn Glu Asn Ser Asn Ile Gln Val Leu Ser Glu Arg Ser Ala
260 265 270
10 Thr Glu Val Asp Asn Asn Phe Ser Lys Pro Pro Phe Phe Pro
275 280 285
Gly Ala Pro Pro Thr His Leu Pro Pro Phe Leu Pro Pro
290 295 300
15 Pro Thr Val Ser Thr Ala Pro Pro Leu Ile Pro Pro Gly Phe Pro
305 310 315 320
Pro Pro Pro Gly Ala Pro Pro Ser Leu Ile Pro Thr Ile Glu Ser
325 330 335
Gly His Ser Ser Gly Tyr Asp Ser Arg Ser Ala Arg Ala Phe Pro Tyr
340 345 350
25 Gly Asn Val Ala Phe Pro His Leu Pro Gly Ser Ala Pro Ser Trp Pro
355 360 365
Ser Leu Val Asp Thr Ser Lys Gln Trp Asp Tyr Tyr Ala Arg Arg Glu
370 375 380
30 Lys Asp Arg Asp Arg Glu Arg Asp Arg Glu Arg Asp Arg Asp
385 390 395 400
Arg Asp Arg Glu Arg Glu Arg Thr Arg Glu Arg Glu Arg Glu Arg Asp
405 410 415
His Ser Pro Thr Pro Ser Val Phe Asn Ser Asp Glu Arg Tyr Arg
420 425 430
40 Tyr Arg Glu Tyr Ala Glu Arg Gly Tyr Glu Arg His Arg Ala Ser Arg
435 440 445
Glu Lys Glu Glu Arg His Arg Glu Arg Arg His Arg Glu Lys Glu Glu
450 455 460
45 Thr Arg His Lys Ser Ser Arg Ser Asn Ser Arg Arg Arg His Glu Ser
465 470 475 480
Glu Glu Gly Asp Ser His Arg Arg His Lys His Lys Lys Ser Lys Arg
485 490 495
50 Ser Lys Glu Gly Lys Glu Ala Gly Ser Glu Pro Ala Pro Glu Gln Glu
500 505 510
55 Ser Thr Glu Ala Thr Pro Ala Glu Xaa
515 520
60 Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe
65 70 75 80

(2) INFORMATION FOR SEQ ID NO: 386:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 137 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 386:
Met Asn Ser Arg Gly Ile Trp Leu Ala Tyr Ile Ile Leu Val Gly Leu
1 5 10 15
Leu His Met Val Leu Leu Ser Ile Pro Phe Phe Ser Ile Pro Val Val
20 25 30
15 Trp Thr Leu Thr Asn Val Ile His Asn Leu Ala Thr Tyr Val Phe Leu
35 40 45
His Thr Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala
50 55 60
20 Arg Leu Leu Thr His Trp Glu Gln Met Asp Tyr Gly Leu Gln Phe Thr
65 70 75 80
Ser Ser Arg Lys Phe Leu Ser Ile Ser Pro Ile Val Leu Tyr Leu Leu
85 90 95
Ala Ser Phe Tyr Thr Lys Tyr Asp Ala Ala His Phe Leu Ile Asn Thr
100 105 110
30 Ala Ser Leu Leu Ser Val Leu Leu Pro Lys Leu Pro Gln Phe His Gly
115 120 125
Val Arg Val Phe Gly Ile Asn Lys Tyr
130 135
35
(2) INFORMATION FOR SEQ ID NO: 387:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 387:
Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu
1 5 10 15
Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg
20 25 30
40 Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe
35 40 45
55 Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Lys Cys
50 55 60
Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe
65 70 75 80

Val Phe Leu Gly Ile Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu 85 95
 5 Leu Val Ala Leu Ala Val Phe Gly Ala Cys Tyr Ile Leu Tyr Leu 100 110
 Arg Thr Leu Gly Ser Lys Leu Val Leu Phe Gly Arg Gly Val Ser Pro 115 125
 10 Ala His Gln Tyr Ala Leu Ala Gly Cys Ile Ser Phe Pro Phe Thr 130 140
 Leu Ala Gly Ala Gly Ser Ala Val Phe Thr Val Leu Gly Ala Thr Leu 145 155
 15 Val Val Ile Gly Ser His Ala Phe His Gln Ile Glu Ala Val Asp 160 175
 Gly Glu Glu Leu Gln Met Glu Pro Val Xaa 180 185

(2) INFORMATION FOR SEQ ID NO: 388:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

Met
1

(2) INFORMATION FOR SEQ ID NO: 389:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 299 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

Met Leu Ser Ile Phe Tyr Phe Ala Ile Pro Val Gly Ser Gly Leu Gly 1 15
 Tyr Ile Ala Gly Ser Lys Val Lys Asp Met Ala Gly Asp Trp His Trp 20 30
 Ala Leu Arg Val Thr Pro Gly Leu Gly Val Val Ala Val Leu Leu Leu 35 45
 Phe Leu Val Val Arg Glu Pro Pro Arg Gly Ala Val Glu Arg His Ser 50 60
 Asp Leu Pro Pro Leu Asn Pro Thr Ser Trp Trp Ala Asp Leu Arg Ala 65 75
 60 Leu Ala Arg Asn Pro Ser Phe Val Leu Ser Ser Leu Gly Phe Thr Ala 80

Val Ala Phe Val Thr Gly Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu 85 95
 5 Leu Arg Ser Arg Val Val Leu Gly Glu Thr Pro Pro Cys Leu Pro Gly 100 110
 Asp Ser Cys Ser Ser Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys 115 125
 Leu Thr Gly Val Leu Gly Val Gly Leu Gly Val Glu Ile Ser Arg Arg 130 140
 10 Leu Arg His Ser Asn Pro Arg Ala Asp Pro Leu Val Cys Ala Thr Gly 145 155
 Leu Leu Gly Ser Ala Pro Phe Leu Ser Leu Ala Cys Ala Arg 160 175
 20 Gly Ser Ile Val Ala Thr Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu 180 190
 Leu Ser Met Asn Trp Ala Ile Val Ala Asp Ile Leu Leu Tyr Val Val 195 205
 25 Ile Pro Thr Arg Arg Ser Thr Ala Glu Ala Phe Gln Ile Val Leu Ser 210 220
 His Leu Leu Gly Asp Ala Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser 225 235
 30 Asp Arg Leu Arg Arg Asn Trp Pro Ser Phe Leu Ser Glu Phe Arg 240 255
 Ala Leu Gln Phe Ser Leu Met Leu Cys Ala Phe Val Gly Ala Leu Gly 260 270
 35 Gly Ala Leu Pro Gly His Arg His Leu His Xaa 275 285
 (2) INFORMATION FOR SEQ ID NO: 390:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

Met Gly Pro Gln Gly Trp Val Arg Pro Leu Lys Thr Ala Pro Lys Leu 1 15
 Gly Glu Ala Ile Arg Leu Ile Leu Phe Leu Asn Phe Val Lys Gln Cys 20 30
 Ile Ala Ser Val Asn Leu Cys Ile Leu Arg Leu Asn Ile Thr Pro Leu 35 45

Leu

5

(2) INFORMATION FOR SEQ ID NO: 391:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 391:

Met Tyr Val Asn Tyr Gly Thr Arg Asn Tyr Ser Thr Glu Gly Pro Ala 15

Ala Leu Leu Asp Gln Ala Lys Leu Ser Leu Leu Val Trp Val Leu Cys 25

20 Phe Val Leu Leu Phe Val Cys Phe Cys Gly Leu Ser Tyr Val Val Ile 45

Ala Gln Val Pro Val Gly Leu Leu Cys Ile Thr Glu Xaa 60

25

(2) INFORMATION FOR SEQ ID NO: 392:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 392:

Met Leu Trp Phe Ala Asn Phe Thr Tyr Leu Phe Leu Ser Gln Ser 15

40 Val Ala Phe Val His Ile Ser His Ile Gly Val Arg Gln Val Asn Thr 30

Asn Cys Tyr Phe Ser Arg Lys Ser Tyr Cys Tyr Gly Ile Leu Asn Pro 45

45 Ile Asn Cys Ile Lys Gly Lys Lys Lys Lys Lys Lys Lys Lys Lys 60

Lys Lys Lys Lys Ile Pro Ala Gly Arg Xaa Leu Phe Pro Phe Gly 75

50

(2) INFORMATION FOR SEQ ID NO: 393:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 393:

60

Met Pro Gly Ala Phe Ser Glu Thr Val Ile Asn Asp Leu Leu Ser Leu 15

5 Phe Leu Val Leu Pro Ala Glu Ser Tyr Ser Thr Leu Ser Gly Val 30

Tyr Arg Asn Ala 35

10

(2) INFORMATION FOR SEQ ID NO: 394:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 180 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 394:

20 Met Ala Gln Ser Arg Asp Gly Gly Asn Pro Phe Ala Glu Pro Ser Glu 15

25 Leu Asp Asn Pro Phe Gln Asp Pro Ala Val Ile Gln His Arg Pro Ser 30

Arg Gln Tyr Ala Thr Leu Asp Val Tyr Asn Pro Phe Glu Thr Arg Glu 45

30 Pro Pro Pro Ala Tyr Glu Pro Pro Ala Pro Ala Pro Leu Pro Pro Pro 60

Ser Ala Pro Ser Leu Gln Pro Ser Arg Lys Leu Ser Pro Thr Glu Pro 80

35 Lys Asn Tyr Gly Ser Tyr Ser Thr Gln Ala Ser Ala Ala Ala Thr 95

Ala Glu Leu Leu Lys Lys Gln Glu Glu Leu Asn Arg Lys Ala Glu Glu 110

40 Leu Asp Arg Arg Ser Glu Ser Cys Ser Met Leu Pro Trp Xaa Ala Gln 125

Leu Leu Asp Arg Thr Ile Gly Pro Leu Tyr Leu Leu Phe Val Gln Phe 140

45 Ser Pro Ala Phe Ser Arg Thr Ser Pro Trp Arg Ser Pro Lys Asn Phe 155

50 Arg Arg Leu Tyr Pro Cys Thr Thr Ser Gly Cys Ala Ala Arg Trp 175

Xaa Phe Ser Xaa 180

55

(2) INFORMATION FOR SEQ ID NO: 395:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 395:
- 5 Met Pro Thr Pro Cys Thr Ser Leu Pro Ser Cys Cys Gln His Arg Ser 15
1 5 10
Ile Thr Met Thr Leu 20
- (2) INFORMATION FOR SEQ ID NO: 396:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 396:
- 20 Met Pro Leu Phe Ile Pro Leu Ile Phe Phe Leu Ser Leu Leu His Cys 15
1 5 10
Gln Ser Lys His Pro Ile Gln Met Ser Leu Cys Met Cys Val Asn Ile 25
20 25 30
Ser Leu Val Trp Ser Pro Val Arg Trp Ile Phe Gly Ser Lys Gly Leu 45
35 40 45
Phe Ser Val His Leu Gln Ser Ser Gln Arg Pro Ser 50
55 60
- (2) INFORMATION FOR SEQ ID NO: 397:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 152 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 397:
- 45 Met Ala Gly Pro Arg Pro Xaa Trp Arg Asp Gln Leu Leu Phe Met Ser 15
1 5 10
Ile Ile Val Leu Val Ile Val Ile Cys Leu Met Leu Tyr Ala Leu 20 25 30
50 Leu Trp Glu Ala Gly Asn Leu Thr Asp Leu Pro Asn Leu Arg Ile Gly 35 40 45
Phe Tyr Asn Phe Cys Leu Trp Asn Glu Asp Thr Ser Thr Leu Gln Cys 50 55 60
His Gln Phe Pro Glu Leu Glu Ala Leu Gly Val Pro Arg Val Gly Leu 65 70 75
Gly Leu Ala Arg Leu Gly Val Tyr Gly Ser Leu Val Leu Thr Leu Phe 80 85 90 95

- Ala Pro Gln Pro Leu Leu Ala Gln Cys Asn Xaa Asp Glu Arg Ala 100 105 110
5 Trp Arg Leu Ala Val Gly Phe Leu Ala Val Ser Ser Val Leu Leu Ala 115 120 125
Gly Gly Leu Gly Leu Phe Leu Ser Tyr Val Trp Asn Gly Ser Xaa Ser 130 135 140
10 Pro Ser Arg Gly Leu Gly Phe Xaa 145 150
- (2) INFORMATION FOR SEQ ID NO: 398:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 480 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 398:
- 15 Met Ser Asp Gly Phe Asp Arg Ala Pro Gly Ala Gly Arg Gly Arg Xaa 5 10 15
25 Arg Gly Leu Gly Arg Gly Gly Gly Pro Xaa Gly Gly Gly Phe Pro 20 25 30
30 Xaa Gly Xaa Xaa Pro Ala Glu Arg Xaa Arg His Gln Pro Pro Gln Pro 35 40 45
Lys Ala Pro Gly Phe Leu Gln Pro Xaa Pro Leu Arg Gln Pro Arg Thr 50 55 60
Thr Pro Pro Gly Ala Gln Cys Glu Val Pro Ala Ser Pro Gln Arg 65 70 75 80
Pro Ser Arg Pro Gly Ala Leu Pro Glu Gln Thr Arg Pro Leu Arg Ala 85 90 95
Pro Pro Ser Ser Gln Asp Lys Ile Pro Gln Gln Asn Ser Glu Ser Ala 100 105 110
45 Met Ala Lys Pro Gln Val Val Val Ala Pro Val Leu Met Ser Lys Leu 115 120 125
Ser Val Asn Ala Pro Glu Phe Tyr Pro Ser Gly Tyr Ser Ser Tyr 130 135 140
50 Thr Glu Ser Tyr Glu Asp Gly Cys Glu Asp Tyr Pro Thr Leu Ser Glu 145 150 155 160
Tyr Val Gln Asp Phe Leu Asn His Leu Thr Glu Gln Pro Gly Ser Phe 165 170 175
80 Glu Thr Glu Ile Glu Gln Phe Ala Glu Thr Leu Asn Gly Cys Val Thr 180 185 190
60 Thr Asp Asp Ala Leu Gln Glu Leu Val Glu Leu Ile Tyr Gln Gln Ala 195

195 200 205
 Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr 210 220
 5 Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln 225 235 240
 Leu Leu Gln Arg Cys Arg Thr Gly Thr Glu Val Lys Asp Gln Ala 245 250 255
 10 Ala Lys Gly Asp Glu Val Thr Arg Lys Arg Phe His Ala Phe Val Leu 260 265 270
 15 Phe Leu Gly Glu Leu Tyr Leu Asn Leu Glu Ile Lys Gly Thr Asn Gly 275 280 285
 Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Glu Leu Leu 290 295 300
 20 Asn Ala Leu Phe Ser Asn Pro Met Asp Asn Asn Leu Ile Cys Ala Val 305 310 315 320
 Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Glu Asp Ala Trp Lys Glu 325 330 335
 25 Lys Gly Lys Met Asp Met Glu Glu Ile Ile Gln Arg Ile Glu Asn Val 340 345 350
 30 Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Leu Lys 355 360 365
 Leu Val Glu Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser 370 375 380
 35 Thr Tyr Arg Glu Ala Thr Pro Glu Asn Asp Pro Asn Tyr Phe Met Asn 385 390 395 400
 Glu Pro Thr Phe Tyr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp 405 410 415
 40 Pro Asp Tyr Gln Glu Lys Tyr Gln Glu Leu Leu Glu Arg Glu Asp Phe 420 425 430
 45 Phe Pro Asp Tyr Glu Glu Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp 435 440 445
 Pro Tyr Leu Asp Asp Ile Asp Asp Glu Met Asp Pro Glu Ile Glu Glu 450 455 460
 50 Ala Tyr Glu Lys Phe Cys Leu Glu Ser Glu Arg Lys Arg Lys Gln Xaa 465 470 475 480
 55
 60 (2) INFORMATION FOR SEQ ID NO: 399:

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 423 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (XI) SEQUENCE DESCRIPTION: SEQ ID NO: 399:
 5 Met Glu Pro Lys Thr Ile Thr Asp Ala Leu Ala Ser Ser Ile Ile Lys 15
 10 Ser Val Leu Pro Asn Phe Leu Pro Tyr Asn Val Met Leu Tyr Ser Asp 20 25 30
 Ala Pro Val Ser Glu Leu Ser Leu Glu Leu Leu Leu Gln Val Val 35 40 45
 15 Leu Pro Ala Leu Leu Glu Gln Gly His Thr Arg Gln Trp Leu Lys Gly 50 55 60
 20 Leu Val Arg Ala Trp Thr Val Thr Ala Gly Tyr Leu Leu Asp Leu His 65 70 75 80
 Ser Tyr Leu Leu Gly Asp Gln Glu Asn Glu Asn Ser Ala Asn Gln 85 90 95
 25 Gln Val Asn Asn Gln His Ala Arg Asn Asn Asn Ala Ile Pro Val 100 105 110
 Val Gly Glu Gly Leu His Ala Ala His Gln Ala Ile Leu Gln Gln Gly 115 120 125
 30 Gly Pro Val Gly Phe Gln Xaa Tyr Arg Arg Pro Leu Asn Phe Pro Leu 130 135 140
 35 Arg Ile Phe Leu Leu Ile Val Phe Met Cys Ile Thr Leu Leu Ile Ala 145 150 155 160
 Ser Leu Ile Cys Leu Thr Leu Pro Val Phe Ala Gly Arg Trp Leu Met 165 170 175
 40 Ser Phe Trp Thr Gly Thr Ala Lys Ile His Glu Leu Tyr Thr Ala Ala 180 185 190
 Cys Gly Leu Tyr Val Cys Trp Leu Thr Ile Arg Ala Val Thr Val Met 195 200 205
 45 Val Ala Trp Met Pro Gln Gly Arg Val Ile Phe Gln Lys Val Lys 210 215 220
 50 Glu Trp Ser Leu Met Ile Met Lys Thr Leu Ile Val Ala Val Leu Leu 225 230 235 240
 Ala Gly Val Val Pro Leu Leu Leu Gly Leu Leu Phe Glu Leu Val Ile 245 250 255
 55 Val Ala Pro Leu Arg Val Pro Leu Asp Gln Thr Pro Leu Phe Tyr Pro 260 265 270
 Trp Gln Asp Trp Ala Leu Gly Val Leu His Ala Lys Ile Ile Ala Ala 275 280 285
 60

Ile Thr Leu Met Gly Pro Gln Trp Trp Leu Lys Thr Val Ile Glu Gln
290 295 300

Val Tyr Ala Asn Gly Ile Arg Asn Ile Asp Leu His Tyr Ile Val Arg
305 310 315 320

Lys Leu Ala Ala Pro Val Ile Ser Val Leu Leu Leu Ser Leu Cys Val
325 330 335

10 Pro Tyr Val Ile Ala Ser Gly Val Val Pro Leu Leu Gly Val Thr Ala
340 345 350

Glu Met Gln Asn Leu Val His Arg Arg Ile Tyr Pro Phe Leu Leu Met
355 360 365

15 Val Val Val Leu Met Ala Ile Leu Ser Phe Gln Val Arg Gln Phe Lys
370 375 380

20 Arg Leu Tyr Glu His Ile Lys Asn Asp Lys Tyr Leu Val Gly Gln Arg
385 390 395 400

Leu Val Asn Tyr Glu Arg Lys Ser Gly Lys Gln Gly Ser Ser Pro Pro
405 410 415

25 Pro Pro Gln Ser Ser Gln Glu
420

30 (2) INFORMATION FOR SEQ ID NO: 400:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 400:

Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met
1 5 10 15

40 Leu Ile Val Ser Val Leu Ala Leu Ile Pro Glu Thr Thr Thr Leu Thr
20 25 30

45 Val Gly Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Cys Leu Ala
35 40 45

Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro
50 55 60

50 Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu Xaa
65 70 75

55 (2) INFORMATION FOR SEQ ID NO: 401:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 401:

Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser
1 5 10 15

5 His Cys Trp Gly Leu Pro Leu His Val Ala Pro Leu Cys Arg Gly His
20 25 30

10 Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala Trp
35 40 45

Asn Arg Asn Leu Ala Asn Gln Arg His Phe Cys Pro Ser Ile Phe
50 55 60

15 His Thr Cys Pro Thr Val Leu Phe Xaa
65 70

20 (2) INFORMATION FOR SEQ ID NO: 402:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 402:

Ala Arg Thr Ile Leu Val Leu Tyr Leu Ser Leu Gln Arg Leu Glu Asn
1 5 10 15

30 Leu Ala Tyr His
20

35 (2) INFORMATION FOR SEQ ID NO: 403:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 403:

Met Pro Leu Pro Ser Val Pro Ile Leu Gly Ile Phe Ser Phe Leu Ile
1 5 10 15

45 Pro Ser Ser Gln Gly Val Ser Tyr Thr Lys Leu Pro Ile Ser Ser Pro
20 25 30

50 Gln Tyr Ser Pro Phe Val Asn Asp His Phe Ser Phe Leu Asn Pro Phe
35 40 45

Pro Val Gln Ile His Thr Gly Phe Ala Arg Val Gly Ser Tyr Met Gln
50 55 60

55 Met Pro Leu Val His Leu Cys Leu Leu Gln Thr Ser Leu Met Lys Asn
65 70 75 80

60 Ser Gly Val Gln Gln Gly Ser
85

- 5 (2) INFORMATION FOR SEQ ID NO: 404:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 92 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 404:
 Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile Ile
 1 5 10 15
 Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala Val
 20 25 30
 Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile Thr
 35 40 45
 Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val Tyr
 50 55 60
 Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg Asp
 65 70 75 80
 Gln Thr Lys Ser Ile Val Glu Lys Ile Pro Ser Lys
 85 90
- 30 (2) INFORMATION FOR SEQ ID NO: 405:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 405:
 Met Ala Cys Ser Cys Leu Met Ile Gln Ser Phe Ser Thr Ser Ala Leu
 1 5 10 15
 Val Leu Phe Tyr Gly
 20
- 45 (2) INFORMATION FOR SEQ ID NO: 406:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 174 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 406:
 Met Glu Glu Gly Gly Asn Leu Gly Leu Ile Lys Met Val His Leu
 1 5 10 15
 Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp Val Thr Phe Val
 20 25 30

- Ser Gly Phe Pro Ala Phe Pro Lys Pro Ser Pro Thr Tyr Leu Arg Thr
 35 40 45
 Ser Ala Glu Gln Thr Leu Pro Leu Leu Pro His Leu His Gly Leu
 50 55 60
 Cys Leu His Gln Pro Leu His Leu Gly Phe Thr Ala Cys Leu Gly Ser
 65 70 75 80
 Ala His Ile Leu Gly Gly Gln Pro Ala Leu Pro Ala Val Pro Glu Pro
 85 90 95
 Tyr Ala Gly His Cys Gln Arg Pro Leu Ala Gly Thr Pro His His Ser
 100 105 110 115
 Cys His Val Gly Pro Ala Asn Arg Gly Arg Arg Ser Glu Ala Trp Val
 120 125 130 135 140
 Gly Arg Tyr Gln Ala Ala Asn Arg Phe Pro Ile Leu Asn Ala Xaa Cys
 145 150 155 160
 Thr Met Gly Cys Pro Leu Phe Ala Ile Trp Ala Ala Ser Xaa
 165 170
- 30 (2) INFORMATION FOR SEQ ID NO: 407:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 64 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 407:
 Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu
 1 5 10 15
 Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His
 20 25 30
 Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys
 35 40 45
 Thr Glu Gly Ile Pro Cys Gln Arg Tyr Gln Asn Gly Leu His Ile Xaa
 50 55 60
- 55 (2) INFORMATION FOR SEQ ID NO: 408:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 280 amino acids
 (B) TYPE: amino acid
- 60

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 408:

Met Glu Ala Val Val Asn Leu Tyr Gln Glu Val Met Lys His Ala Asp 15
1 5 10
Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser Pro Leu Leu Met Thr 30
20 25
Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu Ser Leu Gly Pro Arg 45
35 40
Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg Gly Phe Met Ile Val 60
50 55
Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr Ile Val Tyr Glu Phe 80
65 70 75
Leu Met Ser Gly Trp Leu Ser Thr Tyr Trp Arg Cys Asp Pro Val 95
85 90
Asp Tyr Ser Asn Ser Pro Glu Ala Leu Arg Met Val Arg Val Ala Trp 110
100 105
Leu Phe Leu Phe Ser Lys Phe Ile Glu Leu Met Asp Thr Val Ile Phe 125
115 120
Ile Leu Arg Lys Lys Asp Gly Gln Val Thr Phe Leu His Val Phe His 140
130 135
His Ser Val Leu Pro Trp Ser Trp Trp Trp Gly Val Lys Ile Ala Pro 160
145 150 155
Gly Gly Met Gly Ser Phe His Ala Met Ile Asn Ser Ser Val His Val 175
165 170
Ile Met Tyr Leu Tyr Tyr Gly Leu Ser Ala Phe Gly Pro Val Ala Gln 190
180 185
Pro Tyr Leu Trp Trp Lys Lys His Met Thr Ala Ile Gln Leu Ile Gln 205
195 200
Phe Val Leu Val Ser Leu His Ile Ser Gln Tyr Tyr Phe Met Ser Ser 220
210 215
Cys Asn Tyr Gln Tyr Pro Val Ile Ile His Leu Ile Trp Met Tyr Gly 240
225 230 235
Thr Ile Phe Phe Met Leu Phe Ser Asn Phe Trp Tyr His Ser Tyr Thr 255
245 250
Lys Gly Lys Arg Leu Pro Arg Ala Leu Gln Gln Asn Gly Ala Pro Gly 270
260 265
55 Ile Ala Lys Val Lys Ala Asn Xaa 280
275

(2) INFORMATION FOR SEQ ID NO: 409:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 284 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 409:

Met Xaa Leu Trp Pro Gln Thr Cys Ser Gly Lys Phe Asp Gly Thr Leu 15
1 5 10
Ala Phe Ser Ile His Xaa Leu Ala Val Ile Leu Gly Asp Gln Leu Thr 30
20 25
Ala Ala Asp Leu Val Pro Ile Phe Asn Gly Phe Leu Lys Asp Leu Asp 45
35 40
Glu Val Arg Ile Gly Val Leu Lys His Leu His Asp Phe Leu Lys Leu 60
50 55
Leu His Ile Asp Lys Arg Arg Glu Tyr Leu Tyr Gln Leu Gln Glu Phe 80
65 70 75
Leu Val Thr Asp Asn Ser Arg Asn Trp Arg Phe Arg Ala Glu Leu Ala 95
85 90
Glu Gln Leu Ile Leu Leu Glu Leu Tyr Ser Pro Arg Asp Val Tyr 110
100 105
Asp Tyr Leu Arg Pro Ile Ala Leu Asn Leu Cys Ala Asp Lys Val Ser 125
115 120
Ser Val Arg Trp Ile Ser Tyr Lys Leu Val Ser Glu Met Val Lys Lys 140
130 135 140
Leu His Ala Ala Thr Pro Thr Phe Gly Val Asp Leu Ile Asn Glu 160
145 150 155
Leu Val Glu Asn Phe Gly Arg Cys Pro Lys Trp Ser Gly Arg Gln Ala 175
165 170
Phe Val Phe Val Cys Gln Thr Val Ile Glu Asp Asp Cys Leu Pro Met 190
180 185
Asp Gln Phe Ala Val His Leu Met Pro His Leu Leu Thr Leu Ala Asn 205
195 200
Asp Arg Val Pro Asn Val Arg Val Leu Leu Ala Lys Thr Leu Arg Gln 220
210 215
Thr Leu Leu Glu Lys Asp Tyr Phe Leu Ala Ser Ala Ser Cys His Gln 240
225 230 235
Glu Ala Val Glu Gln Thr Ile Met Ala Leu Gln Met Asp Arg Asp Ser 255
245 250
Asp Val Lys Tyr Phe Ala Ser Ile His Pro Ala Ser Thr Lys Ile Ser 270
260 265
60 Glu Asp Ala Met Ser Thr Ala Ser Ser Thr Tyr Xaa 280
275

600

601

(2) INFORMATION FOR SEQ ID NO: 410:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 187 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 410:

Met Leu Phe Leu Phe Phe Val Ile Ile Phe Leu Val Phe Leu Ile 15
1 5 10
Leu Ile Ile Gln Phe Ser Lys Pro Leu Thr Asn Pro His Pro Pro Ala 30
20
Gly Xaa Ser Asp Arg Arg Arg Tyr Ser Ser Tyr Arg Ser His Asp 45
35
His Tyr Gln Arg Gln Arg Val Leu Gln Lys Glu Arg Ala Ile Glu Glu 60
50
Arg Arg Val Phe Ile Gly Lys Ile Pro Gly Arg Met Thr Arg Ser 80
65
Glu Leu Lys Gln Arg Phe Ser Val Phe Gly Glu Ile Glu Glu Cys Thr 95
85
Ile His Phe Arg Val Gln Gly Asp Asn Tyr Gly Phe Val Thr Tyr Arg 110
100
Tyr Ala Glu Glu Ala Phe Ala Ile Glu Ser Gly His Lys Leu Arg 125
115
Gln Ala Asp Glu Gln Pro Phe Asp Leu Cys Phe Gly Gly Arg Arg Xaa 140
130
Xaa Cys Lys Arg Ser Tyr Ser Asp Leu Asp Ser Asn Arg Glu Asp Phe 155
145
Asp Pro Ala Pro Val Lys Ser Lys Phe Asp Ser Leu Asp Phe Asp Thr 175
165
Leu Leu Lys Gln Ala Gln Lys Asn Leu Arg Arg 185
180

(2) INFORMATION FOR SEQ ID NO: 411:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 237 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

Met Lys Leu Pro Gly Lys Phe Arg Arg Ala His Gln Gly Asn Leu Glu 15
1 5 10

Ser Gln Leu Thr Ser Glu Ser Tyr Tyr Lys Glu Thr Leu Ser Val Pro 30
20
Thr Val Glu His Ile Ile Gln Glu Leu Lys Asp Ile Phe Ser Glu Gln 45
35
His Leu Lys Ala Leu Lys Cys Leu Ser Leu Val Pro Ser Val Met Gly 60
50
Gln Leu Lys Phe Asn Thr Ser Glu Glu His Ala Asp Met Tyr Arg 80
65
Ser Asp Leu Pro Asn Pro Asp Thr Leu Ser Ala Glu Leu His Cys Trp 95
85
Arg Ile Lys Trp Lys His Arg Gly Lys Asp Ile Glu Leu Pro Ser Thr 110
100
Ile Tyr Glu Ala Leu His Leu Pro Asp Ile Lys Phe Phe Pro Asn Val 125
115
Tyr Ala Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu 140
130
Asn Glu Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg 160
145
Asn Thr Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile 175
165
Asn Phe Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile 190
180
Lys Leu Tyr Thr Xaa Xaa Ser Xaa Leu Xaa Thr Xaa Xaa Ser Xaa Xaa 205
195
Val Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa 220
210
Asp Xaa Xaa Xaa Arg Glu Lys Ala Val Arg Cys Met Xaa 235
225
Met Lys Pro Met Ala Val Val Ala Ser Thr Val Leu Gly Leu Val Gln 15
1 5 10
Asn Met Arg Ala Phe Gly Gly Ile Leu Val Val Val Tyr Tyr Val Phe 30
20
Ala Ile Ile Gly Ile Asn Leu Phe Arg Gly Val Ile Val Ala Leu Pro 45
35

Gly Asn Ser Ser Leu Ala Pro Ala Asn Gly Ser Ala Pro Cys Gly Ser
50 55 60

5 Phe Glu Cln Leu Glu Tyr Trp Ala Asn Asn Phe Asp Asp Phe Ala Ala
65 70 75 80

Ala Leu Val Thr Leu Trp Asn Leu Met Val Val Asn Asn Trp Gln Val
85 90 95

10 Phe Leu Asp Ala Tyr Arg Arg Tyr Ser Gly Pro Trp Ser Lys Ile Tyr
100 105 110

15 Phe Val Leu Trp Trp Leu Val Ser Ser Val Ile Trp Val Asn Leu Phe
115 120 125

Leu Ala Leu Ile Leu Glu Asn Phe Leu His Lys Lys Trp Asp Pro Arg Ser
130 135 140

20 His Leu Gln Pro Leu Ala Gly Thr Pro Glu Ala Thr Tyr Gln Met Thr
145 150 155 160

Val Glu Leu Leu Phe Arg Asp Ile Leu Glu Glu Pro Gly Glu Asp Glu
165 170 175

25 Leu Thr Glu Arg Leu Ser Gln His Pro His Leu Trp Leu Cys Arg Xaa
180 185 190

30

(2) INFORMATION FOR SEQ ID NO: 413:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 413:

Asn Val Val Val Val Ala Phe Gly Leu Ile Leu Ile Ile Glu Ser Leu
1 5 10 15

45

Gly Glu Cln Cys Pro
20

(2) INFORMATION FOR SEQ ID NO: 414:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 414:

Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys Lys
1 5 10 15

60

Glu Cln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala
20 25 30

5 Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Ser Leu Leu Lys Met Arg Arg
35 40 45

Ala Pro Xaa
50

10

(2) INFORMATION FOR SEQ ID NO: 415:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 415:

20 Met Leu Ile Ile Ser Leu Arg Pro Gln Phe Pro Ser Leu Ile Val Gln
1 5 10 15

Leu Glu Cys Ser Val Leu Phe Leu Pro Ile Ser Leu Asn Leu Leu Leu
20 25 30

25

30

(2) INFORMATION FOR SEQ ID NO: 416:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 163 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 416:

Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu
1 5 10 15

Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu
20 25 30

45 Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp
35 40 45

Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr
50 55 60

50 Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu
65 70 75 80

Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu
85 90 95

55 Glu Ile Arg Gly Gly Leu Lys Tyr Cys Asn Leu Leu Val Arg
100 105 110

60 Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met

115 120 125
Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Gln Arg
130 135 140
5 Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Ile Gly Arg Pro
145 150 155 160
Gln Val Xaa
10

(2) INFORMATION FOR SEQ ID NO: 417:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 417:
Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe
1 5 10 15
25 Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly
20 25 30
Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Phe Tyr Ser Ser
35 40 45
30 Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr
50 55 60
Gly Ala Ser Tyr His Ser Met Ser Ser Met Ala Arg Ala Ala Phe Ser
65 70 75 80
Glu Asp Gly Ala Leu Met Asp Gly Gly Met Asp Leu Asn Met Glu Gln
85 90 95
40 Gly Met Ala Glu His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val
100 105 110
Gln Val Leu Ser Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu
115 120 125
45 Ala Pro Gly Arg Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro
130 135 140
Trp Phe Thr Ala Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys
145 150 155 160
Arg Gln Arg Arg Gln Glu Arg Arg Gln Met Lys Arg Leu Xaa
165 170
50

(2) INFORMATION FOR SEQ ID NO: 418:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 418:
5 Met Glu Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Val Ala Met Met
1 5 10 15
Arg Pro Phe Tyr Leu Leu Pro Val Leu Cys Thr Gln Ala Leu Arg
20 25 30
10 Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Cys Cys Leu
35 40 45
Ala Xaa
15 50

(2) INFORMATION FOR SEQ ID NO: 419:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 120 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 419:
Met Leu Gly Lys Gly Gly Arg Ala Gly Leu Leu Arg Tyr Arg Leu
1 5 10 15
30 Leu Tyr Phe Thr Leu Val Val Gly Glu Gly Glu Pro Gly Glu Asn Lys
20 25 30
Val Thr Ile Pro Phe Phe Glu Thr Gly Lys Lys Ile Ile Phe Cys Ser
35 40 45
35 Val Lys Met Val Glu Asn Ser Asn Val Pro Ser His Lys Gly Pro Val
50 55 60
Pro Leu Arg Ser Glu Gln Trp Glu Lys Ile Ser Glu Thr Leu Gly
65 70 75 80
Glu Gly Lys Ile Gly Phe Leu Leu Ile Gly Arg Cys Ser Ser Gly Xaa
85 90 95
45 Gly Gly Leu Cys Phe Cys Trp Asp Val Leu Cys Cys Met Tyr Ala Tyr
100 105 110
Met Asp Arg Ser Leu Leu Ser Leu
115 120
50

(2) INFORMATION FOR SEQ ID NO: 420:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 159 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 420:
55

60

Met Thr His Leu Leu Thr Ala Thr Val Thr Pro Ser Glu Gln Asn
1 5 10 15
Ser Ser Arg Glu Pro Gly Trp Glu Thr Ala Met Ala Lys Asp Ile Leu
20 25 30
Gly Glu Ala Gly Leu His Phe Asp Glu Leu Asn Lys Leu Arg Val Leu
35 40 45
Asp Pro Glu Val Thr Gln Gln Thr Ile Glu Leu Lys Glu Glu Cys Lys
50 55 60
Asp Phe Val Asp Lys Ile Gly Gln Phe Gln Lys Ile Val Gly Gly Leu
65 70 75 80
Ile Glu Leu Val Asp Gln Leu Ala Lys Glu Ala Glu Asn Glu Lys Met
85 90 95
Lys Ala Ile Gly Ala Arg Asn Leu Leu Lys Ser Ile Ala Lys Gln Arg
100 105 110
Glu Ala Gln Gln Gln Leu Gln Ala Leu Ile Ala Glu Lys Lys Met
115 120 125
Gln Leu Glu Arg Tyr Arg Val Glu Tyr Glu Ala Leu Cys Lys Val Glu
130 135 140
Ala Glu Gln Asn Glu Phe Ile Asp Gln Phe Ile Phe Gln Lys Xaa
145 150 155
(2) INFORMATION FOR SEQ ID NO: 421:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 154 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 421:
Met Asn Val Gly Val Ala His Ser Glu Val Asn Pro Asn Thr Arg Val
1 5 10 15
Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu
20 25 30
Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala
35 40 45
Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu
50 55 60
His Ala Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala
65 70 75 80
Arg Leu Leu Thr His Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr
85 90 95
Ser Ser Arg Lys Phe Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu
100 105 110

Ala Ser Phe Tyr Thr Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr
115 120 125
5 Ala Ser Leu Leu Ser Val Leu Ile Pro Lys Met Pro Gln Leu His Gly
130 135 140
Val Arg Ile Phe Gly Ile Asn Lys Tyr Xaa
145 150
10
(2) INFORMATION FOR SEQ ID NO: 422:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 204 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 422:
Met Val Cys Gly Gly Phe Ala Cys Ser Lys Asn Cys Leu Cys Ala Leu
1 5 10 15
Asn Leu Leu Tyr Thr Leu Val Ser Leu Leu Ile Ile Gly Ile Ala Ala
20 25 30
Trp Gly Ile Gly Phe Gly Leu Ile Ser Ser Leu Arg Val Val Gly Val
35 40 45
Val Ile Ala Val Gly Ile Phe Leu Phe Leu Ile Ala Leu Val Gly Leu
50 55 60
Ile Gly Ala Val Lys His Gln Val Leu Leu Phe Phe Tyr Met Ile
65 70 75 80
Ile Leu Leu Val Phe Ile Val Gln Phe Ser Val Ser Cys Ala Cys
85 90 95
Leu Ala Leu Asn Gln Glu Gln Gln Gly Gln Leu Leu Glu Val Gly Trp
100 105 110
Asn Asn Thr Ala Ser Ala Ser Arg Asn Asp Ile Gln Arg Asn Leu Asn Cys
115 120 125
Cys Gly Phe Arg Ser Val Asn Pro Asn Asp Thr Cys Leu Ala Ser Cys
130 135 140
Val Lys Ser Asp His Ser Cys Ser Pro Cys Ala Pro Ile Ile Gly Glu
145 150 155 160
Tyr Ala Gly Glu Val Leu Arg Phe Val Gly Ile Gly Leu Phe Phe
165 170 175
Ser Phe Thr Glu Ile Leu Gly Val Trp Leu Thr Tyr Arg Tyr Arg Asn
180 185 190
Gln Lys Asp Pro Arg Ala Asn Pro Ser Ala Phe Leu
195 200
60

(2) INFORMATION FOR SEQ ID NO: 423:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 67 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 423:

10 Met Leu Gln Ser Ile Ile Lys Asn Ile Trp Ile Pro Met Lys Pro Tyr
1 5 10 15

Tyr Thr Lys Val Tyr Gln Glu Ile Trp Ile Gly Met Gly Leu Met Gly
20 25 30

15 Phe Ile Val Tyr Lys Ile Arg Ala Ala Asp Lys Arg Ser Lys Ala Leu
35 40 45

20 Lys Ala Ser Ala Pro Ala Pro Gly His His Asn Gln Ile Tyr Leu Glu
50 55 60

Tyr Met Xaa
65

(2) INFORMATION FOR SEQ ID NO: 424:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 424:

35 Met Leu Gly Val Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val
1 5 10 15

Ala Val Asn Asn Pro Lys Lys Gln Glu
20 25

(2) INFORMATION FOR SEQ ID NO: 425:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 299 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 425:

45 Met Ala Ala Xaa Glu Pro Ala Val Leu Ala Leu Pro Asn Ser Gly Ala
1 5 10 15

Gly Gly Ala Gly Ala Pro Ser Gly Thr Val Pro Val Leu Phe Cys Phe
20 25 30

55 Ser Val Phe Ala Arg Pro Ser Ser Val Pro His Gly Ala Gly Tyr Glu
35 40 45

60 Leu Leu Ile Gln Lys Phe Leu Ser Leu Tyr Gly Asp Gln Ile Asp Met
5 10

50 55 60

His Arg Lys Phe Val Val Gln Leu Phe Ala Glu Glu Trp Gly Gln Tyr
65 70 75 80

5 Val Asp Leu Pro Lys Gly Phe Ala Val Ser Glu Arg Cys Lys Val Arg
85 90 95

10 Leu Val Pro Leu Gln Ile Gln Leu Thr Thr Leu Gly Asn Leu Thr Pro
100 105 110

Ser Ser Thr Val Phe Phe Cys Asp Met Gln Glu Arg Phe Arg Pro
115 120 125

15 Ala Ile Lys Tyr Phe Gly Asp Ile Ile Ser Val Gly Gln Arg Leu
130 135 140

Gln Gly Ala Arg Ile Leu Gly Ile Pro Val Ile Val Thr Glu Gln Tyr
145 150 155 160

20 Pro Lys Gly Leu Gly Ser Thr Val Gln Glu Ile Asp Leu Thr Gly Val
165 170 175

25 Lys Leu Val Leu Pro Lys Thr Lys Phe Ser Met Val Leu Pro Glu Val
180 185 190

Glu Ala Ala Leu Ala Glu Ile Pro Gly Val Arg Ser Val Val Leu Phe
195 200 205

30 Gly Val Glu Thr His Val Cys Ile Gln Gln Thr Ala Leu Glu Leu Val
210 215 220

Gly Arg Gly Val Glu Val His Ile Val Ala Asp Ala Thr Ser Ser Arg
225 230 235 240

35 Ser Met Met Asp Arg Met Phe Ala Leu Glu Arg Leu Ala Xaa Xaa Gly
245 250 255

40 Ile Ile Val Thr Thr Ser Glu Ala Val Leu Leu Gln Leu Val Ala Asp
260 265 270

Lys Asp His Pro Lys Phe Lys Glu Ile Gln Asn Leu Ile Lys Ala Ser
275 280 285

45 Ala Pro Glu Ser Gly Leu Leu Ser Lys Val Xaa
290 295

(2) INFORMATION FOR SEQ ID NO: 426:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 426:

Met Arg Asp Leu Gly Thr Leu Leu Ser Pro Val Cys Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 427:

5

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 427:

10

Met Phe Gly Cys Leu Val Ala Gly Arg Leu Val Gln Thr Ala Ala Gln
1 5 10 15

15

Gln Val Ala Glu Asp Lys Phe Val Phe Asp Leu Pro Asp Tyr Glu Ser
20 25 30Ile Asn His Val Val Val Phe Met Leu Gly Thr Ile Pro Phe Pro Glu
35 40 45

20

Gly Met Gly Gly Ser Val Tyr Phe Ser Tyr Pro Asp Ser Asn Gly Met
50 55 60Pro Val Trp Gln Leu Leu Gly Phe Val Thr Asn Gly Lys Pro Ser Ala
65 70 75 80

25

Ile Phe Lys Ile Ser Gly Leu Lys Ser Gly Glu Gly Ser Gln His Pro
85 90 95

30

Phe Gly Ala Met Asn Ile Val Arg Thr Pro Ser Val Ala Gln Ile Gly
100 105 110Ile Ser Val Glu Leu Asp Ser Met Ala Gln Gln Thr Pro Val Gly
115 120 125

35

Asn Ala Ala Val Ser Val Asp Ser Phe Thr Gln Phe Thr Gln Lys
130 135 140Met Leu Asp Asn Phe Tyr Asn Phe Ala Ser Ser Phe Ala Val Ser Gln
145 150 155 160

40

Ala Gln Met Thr Pro Ser Pro Ser Glu Met Phe Ile Pro Ala Asn Val
165 170 175

45

Val Leu Lys Tyr Tyr Glu Asn Phe Gln Arg Arg Leu Ala Gln Asn Pro
180 185 190Xaa Phe Trp Xaa Thr Xaa
195

50

(2) INFORMATION FOR SEQ ID NO: 428:

55

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 428:

60 Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser

1

5

10

15

Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Leu Trp Lys
20 25 30

5

Asn Arg Gly Gly Val Gly Arg Ser Val Met Ser Ala Val Glu Xaa
35 40 45

10

(2) INFORMATION FOR SEQ ID NO: 429:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 370 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 429:

20

Met Lys Lys Val Glu Glu Lys Arg Val Asp Val Asn Ser Ala Val Ala
1 5 10 15

25

Met Gly Glu Val Ile Leu Ala Val Cys His Pro Asp Cys Ile Thr Thr
20 25 30

25

Ile Lys His Trp Ile Thr Ile Ile Arg Ala Arg Phe Glu Glu Val Leu
35 40 45

30

Thr Trp Ala Lys Gln His Gln Arg Leu Glu Thr Ala Leu Ser Glu
50 55 60

35

Leu Val Ala Asn Ala Glu Leu Leu Glu Glu Leu Ala Trp Ile Gln
65 70 75 80

40

Trp Ala Glu Thr Thr Leu Ile Gln Arg Asp Gln Glu Pro Ile Pro Gln
85 90 95

45

Asn Ile Asp Arg Val Lys Ala Leu Ile Ala Glu His Gln Thr Phe Met
100 105 110

50

Glu Glu Met Thr Arg Lys Gln Pro Asp Val Asp Arg Val Thr Lys Thr
115 120 125

55

Tyr Lys Arg Lys Asn Ile Glu Pro Thr His Ala Pro Phe Ile Glu Lys
130 135 140

60

Ser Arg Ser Gly Gly Arg Lys Ser Leu Ser Gln Pro Thr Pro Pro Pro
145 150 155 160

65

Met Pro Ile Leu Ser Gln Ser Glu Ala Lys Asn Pro Arg Ile Asn Gln
165 170 175

70

Leu Ser Ala Arg Trp Gln Gln Val Trp Leu Leu Ala Leu Glu Arg Gln
180 185 190

75

Arg Lys Leu Asn Asp Ala Leu Asp Arg Leu Glu Leu Lys Glu Phe
195 200 205

80

Ala Asn Phe Asp Phe Asp Val Trp Arg Lys Tyr Met Arg Trp Met
210 215 220 225

Asn His Lys Lys Ser Arg Val Met Asp Phe Phe Arg Arg Ile Asp Lys
225 230 235

Asp Gln Asp Gly Lys Ile Thr Arg Gln Glu Phe Ile Asp Gly Ile Leu
245 250 255

Ala Ser Lys Phe Pro Thr Thr Lys Leu Glu Met Thr Ala Val Ala Asp
260 265 270

Ile Phe Asp Arg Asp Gly Tyr Ile Asp Tyr Tyr Glu Phe Val
275 280 285

Ala Ala Leu His Pro Asn Lys Asp Ala Tyr Arg Pro Thr Thr Asp Ala
290 295 300

Asp Lys Ile Glu Asp Glu Val Thr Arg Gln Val Ala Gln Cys Lys Cys
305 310 315

Ala Lys Arg Phe Gln Val Glu Gln Ile Gly Glu Asn Lys Tyr Arg Phe
325 330 335

Phe Leu Gly Asn Gln Phe Gly Asp Ser Gln Gln Leu Arg Leu Val Arg
340 345 350

Ile Leu Arg Asn Arg Asp Gly Ser Arg Trp Trp Arg Met Asp Gly Leu
355 360 365

Gly Xaa
370

(2) INFORMATION FOR SEQ ID NO: 430:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 430:

Met Asn Val Lys Thr Phe Ser Xaa Asp His Met His Phe Leu Cys Cys
1 5 10 15

Leu Tyr Leu Arg Tyr Val Thr Phe Val Tyr Leu Asn Leu Phe
20 25 30

(2) INFORMATION FOR SEQ ID NO: 431:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 431:

Met Glu Pro His Leu Arg Cys Arg Val Thr Arg Val Arg Gly Ser Leu
1 5 10 15

Gly Asn Thr Gly Arg Trp Leu Leu

20

5 (2) INFORMATION FOR SEQ ID NO: 432:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 432:

Met His Tyr Leu Val Leu Gly Gly Leu Gly Val Phe Leu Phe Phe Ser
1 5 10 15

Cys Phe Val Phe Leu Phe Phe Xaa Phe Ser Phe Ala Phe Phe Pro Phe
20 25 30

Tyr Leu Glu Gly Met Gly Gly Ser Gly Asn Arg Glu Val Gly Gly Gly
35 40 45

Phe Cys Leu Phe Phe
50

(2) INFORMATION FOR SEQ ID NO: 433:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 176 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 433:

Met Val Ser Lys Ala Leu Leu Arg Leu Val Ser Ala Val Asn Arg Arg
1 5 10 15

Arg Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val Ala Ser
20 25 30

Val Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu
35 40 45

Leu Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu
50 55 60

Lys Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val
65 70 75 80

Lys Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Ile Thr Gly Gly
85 90 95

Ala Gly Phe Val Gly Ser His Leu Thr Asp Lys Leu Met Met Asp Gly
100 105 110

His Glu Val Thr Val Val Asp Asn Phe Phe Thr Gly Arg Lys Arg Asn
115 120 125

Val Glu His Trp Ile Gly His Glu Asn Phe Glu Leu Ile Asn His Asp
130 135 140

Val Trp Ser Pro Ser Thr Arg Leu Thr Arg Tyr Thr Ile Trp His
145 150 155

5 Leu Gln Pro Pro Leu Gln Thr Thr Cys Ile Ile Leu Ser Arg His Xaa
165 170 175

10

(2) INFORMATION FOR SEQ ID NO: 434:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 77 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 434:

20 Met Leu Arg Cys Trp Pro Leu Phe Trp Leu Pro Leu Val Ser Pro Phe
1 5 10 15

25 Cys Ser Leu Phe Trp Leu Val Glu Trp Phe Gly Thr Asn Ile Asp
20 25 30

Arg Glu Ser Tyr Asp Ala Ile Gly Gly Pro Ser Trp Met Thr Ala Ser
35 40 45

30 Ser Phe Cys Leu Ser Asn Ser Asn Ile Trp Ser Leu Glu Ile Ser Ser
50 55 60

Gly Ser Thr Val Val His Ser Gln Gln Ala Met Asp
65 70 75

35

(2) INFORMATION FOR SEQ ID NO: 435:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 435:

45 Met Arg Ser Cys Glu Ile Gln Leu Cys Val Trp Leu Leu Val Ser Ser
1 5 10 15

50 His Val Asp Met Val Leu Gly Gly Ser Pro Ser Thr Leu Tyr Met Met
20 25 30

55

(2) INFORMATION FOR SEQ ID NO: 436:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids

60

- (B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 436:

5 Met Val Val Asn Ser Leu Cys Phe Leu Ser Leu Leu Val Ile Leu
1 5 10 15

Glu Leu Ser Thr Asp Ser Ser Ala Arg Leu Leu Tyr His Glu
20 25 30

10

(2) INFORMATION FOR SEQ ID NO: 437:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 437:

20 Met Asp Lys Gln Lys His Leu Glu Val Arg Arg Ser Val Phe Lys Ile
1 5 10 15

25 Gln Gly Lys Ile Ala Phe Ser Leu Met Phe Val Leu Lys Asp Leu Ser
20 25 30

Pro Thr Ile Phe Ser His Ser Ile Leu Leu Leu Pro His His Val
35 40 45

30 Leu Pro Cys Thr Pro Gln Met Val Arg Gly Val Thr Gln Val Leu Arg
50 55 60

Glu Phe Gly Asp Gln
65

35

(2) INFORMATION FOR SEQ ID NO: 438:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 438:

45 Met Pro Leu Cys Phe Phe Ser Phe Leu Cys Cys Trp Val Leu Val Phe
1 5 10 15

Lys Leu Ile
50

(2) INFORMATION FOR SEQ ID NO: 439:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 439:

60

Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg
1 5 10 15

5 Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg
20 25 30

Arg Val Met Val Asn Leu Asn Ile Leu Phe Xaa
35 40

10

(2) INFORMATION FOR SEQ ID NO: 440:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 440:

20 Met Leu Leu Phe Pro Ser Leu Leu Phe Ala Ala Thr Tyr Asn Val Ala
1 5 10 15

25 Asn Pro Ser Arg Leu Ile Leu Tyr Met Ile Ser Ala Gly Ala Asp Ser
20 25 30

Gln

(2) INFORMATION FOR SEQ ID NO: 441:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 441:

40 Met Trp Gln Val Arg Gly Leu Pro Pro Val Pro Leu Leu Thr Met
1 5 10 15

Ser Pro Pro Cys Leu Ser Ser Pro Phe Pro Phe Ile Ser Val Pro
20 25 30

45 Leu Phe Glu Ala Val Pro Ile Ser Val Ser Asp Gln Pro Ser Pro Xaa
35 40 45

50 Leu Thr Thr Leu Leu
50

(2) INFORMATION FOR SEQ ID NO: 442:

55 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 442:

60

Met Ile Thr Ser Val Leu Val Phe Leu Ile Phe Phe Pro Tyr Leu
1 5 10 15

5 Ser Leu Val Thr Leu Leu Gln Ala Arg Asn Leu Trp Val Ile His Arg
20 25 30

10 Ala Ala Leu Cys Glu Ser Gly Leu Phe His Trp Arg Lys Gly Ile Glu
35 40 45

Asn Gln Leu Glu Pro Met Tyr Phe Leu Pro His Gly Thr Leu Phe Leu
50 55 60

15

(2) INFORMATION FOR SEQ ID NO: 443:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 443:

Met Leu Tyr Ser Cys Glu Pro Tyr Leu Ile Ile Leu Asn Ile Tyr Ser
1 5 10 15

30 Gln Lys Ala Phe Tyr Phe Phe Glu Gly Ser Phe Ser Val Cys
20 25 30

Thr Leu

35

(2) INFORMATION FOR SEQ ID NO: 444:

40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 89 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 444:

Met Arg Gln Arg Gln Ala Ala Cys Gln Pro Pro Ser Arg Asn Gly
1 5 10 15

50 Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Phe Leu Val
20 25 30

Lys Leu Leu Phe Ile Pro Trp Trp Ala Ser Leu Leu Ser Ser Pro Leu
35 40 45

55 Asn Leu Leu Leu Val Ser Ile Ser Trp Asp Leu Gly Leu Lys Leu
50 55 60

60 Asn Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr
65 70 75 80

618

Lys Lys Phe Asn Lys Lys Lys Lys Lys
85

5

(2) INFORMATION FOR SEQ ID NO: 445:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 350 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 445:

Met Asp Phe Ile Thr Ser Thr Ala Ile Leu Pro Leu Leu Phe Gly Cys
1 10 15

Leu Gly Val Phe Gly Leu Phe Arg Leu Leu Gln Trp Val Arg Gly Lys
20 25 30

Ala Tyr Leu Arg Asn Ala Val Val Ile Thr Gly Ala Thr Ser Gly
35 40 45

Leu Gly Lys Glu Cys Ala Lys Val Phe Tyr Ala Ala Gly Ala Lys Leu
50 55 60

Val Leu Cys Gly Arg Asn Gly Gly Ala Leu Glu Glu Leu Ile Arg Glu
65 70 75 80

Leu Thr Ala Ser His Ala Thr Lys Val Gln Thr His Lys Pro Tyr Leu
85 90 95

Val Thr Phe Asp Leu Thr Asp Ser Gly Ala Ile Val Ala Ala Ala
100 105 110

Glu Ile Leu Gln Cys Phe Gly Tyr Val Asp Ile Leu Val Asn Ala
115 120 125

Gly Ile Ser Tyr Arg Gly Thr Ile Met Asp Thr Thr Val Asp Val Asp
130 135 140

Lys Arg Val Met Glu Thr Asn Tyr Phe Gly Pro Val Ala Leu Thr Lys
145 150 155 160

Ala Leu Leu Pro Ser Met Ile Lys Arg Arg Gln Gly His Ile Val Ala
165 170 175

Ile Ser Ser Ile Gln Gly Lys Met Ser Ile Pro Phe Arg Ser Ala Tyr
180 185 190

Ala Ala Ser Lys His Ala Thr Gln Ala Phe Phe Asp Cys Leu Arg Ala
195 200 205

Glu Met Glu Gln Tyr Glu Ile Glu Val Thr Val Ile Ser Pro Gly Tyr
210 215 220

Ile His Thr Asn Leu Ser Val Asn Ala Ile Thr Ala Asp Gly Ser Arg
225 230 235 240

Tyr Gly Val Met Asp Thr Thr Thr Ala Gln Gly Arg Ser Pro Val Glu
245 250 255

60

619

Val Ala Gln Asp Val Leu Ala Ala Val Gly Lys Lys Lys Lys Asp Val
260 265 270

Ile Leu Ala Asp Leu Leu Pro Ser Leu Ala Val Tyr Leu Arg Thr Leu
275 280 285

Ala Pro Gly Leu Phe Phe Ser Leu Met Pro Pro Gly Pro Glu Lys Ser
290 295 300

Gly Asn Pro Arg Thr Pro Ser Thr Leu Thr Ser Gln Gly Gln Gly Arg
305 310 315 320

Glu Ala Ala Leu Leu Gly Leu Thr Leu Gln Gly Thr Val Ala Phe
325 330 335

Val Glu Thr Leu Met Glu Ile Cys Leu Thr Ser Gly Lys Asp
340 345 350

20

(2) INFORMATION FOR SEQ ID NO: 446:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 446:

Met Val Phe Leu Pro Arg Gly Val Val Val Ser Gly Gly Ala Ala Cys
1 5 10 15

Leu Trp Leu Thr Phe Ile Leu Glu Thr Glu Val Tyr Leu Asp Leu Ala
20 25 30

Thr Glu Ala Arg Ala His Ser Arg Met Gly Leu Gly Leu Trp Pro Pro
35 40 45

Asn

40

(2) INFORMATION FOR SEQ ID NO: 447:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 447:

Met Ala Ser Ala Glu Leu Asp Tyr Thr Ile Glu Ile Pro Asp Gln Pro
1 5 10 15

Cys Trp Ser Gln Lys Asn Ser Pro Ser Pro Gly Gly Lys Glu Ala Glu
20 25 30

Thr Arg Gln Pro Val Val Ile Leu Leu Gly Trp Gly Gly Cys Lys Asp
35 40 45

60

Lys Asn Leu Ala Lys Tyr Ser Ala Ile Tyr His Lys Arg Gly Cys Ile
50 55 60

Val Ile Arg Tyr Thr Ala Pro Trp His Met Val Phe Phe Ser Glu Ser
65 70 75 80

Leu Gly Ile Pro Ser Leu Arg Val Leu Ala Gln Lys Leu Leu Glu Leu
85 90 95

10 Leu Phe Asp Tyr Glu Ile Glu Lys Glu Pro Leu Leu Phe His Val Phe
100 105 110

Ser Asn Gly Gly Val Met Leu Tyr Arg Tyr Val Leu Glu Leu Leu Gln
115 120 125

15 Thr Arg Arg Phe Cys Arg Leu Arg Val Val Gly Thr Ile Phe Asp Ser
130 135 140

20 Ala Pro Gly Asp Ser Asn Leu Val Gly Ala Leu Arg Ala Leu Ala Ala
145 150 155 160

Ile Leu Glu Arg Arg Ala Ala Met Leu Arg Leu Leu Leu Val Ala
165 170 175

25 Phe Ala Leu Val Val Leu Phe His Val Leu Leu Ala Pro Ile Thr
180 185 190

Ala Xaa Phe His Thr Phe Tyr Asp Arg Leu Gln Asp Ala Gly Ser
195 200 205

30 Arg Trp Pro Glu Leu Tyr Leu Tyr Ser Arg Ala Asp Glu Val Val Leu
210 215 220

35 Ala Arg Asp Ile Glu Arg Met Val Glu Ala Arg Leu Ala Arg Arg Val
225 230 235 240

Leu Ala Arg Ser Val Asp Phe Val Ser Ser Ala His Val Ser His Leu
245 250 255

40 Arg Asp Tyr Pro Thr Tyr Thr Ser Leu Cys Val Asp Phe Met Arg
260 265 270

Asn Cys Val Arg Cys Xaa
275

(2) INFORMATION FOR SEQ ID NO: 448:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 199 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 448:

Met Ser Phe Ile Phe Asp Trp Ile Tyr Ser Gly Phe Ser Ser Val Leu
1 5 10 15

Gln Phe Leu Gly Leu Tyr Lys Lys Thr Gly Lys Leu Val Phe Leu Gly
20 25 30

Leu Asp Asn Ala Gly Lys Thr Thr Leu Leu His Met Leu Lys Asp Asp
35 40 45

5 Arg Leu Gly Gln His Val Pro Thr Leu His Pro Thr Ser Glu Glu Leu
50 55 60

10 Thr Ile Ala Gly Met Thr Phe Thr Thr Phe Asp Leu Gly Gly His Val
65 70 75 80

Gln Ala Arg Arg Val Trp Lys Asn Tyr Leu Pro Ala Ile Asn Gly Ile
85 90 95

15 Val Phe Leu Val Asp Cys Ala Asp His Glu Arg Leu Leu Glu Ser Lys
100 105 110

Glu Glu Leu Asp Ser Leu Met Thr Asp Glu Thr Ile Ala Asn Val Pro
115 120 125

20 Ile Leu Ile Leu Gly Asn Lys Ile Asp Arg Pro Glu Ala Ile Ser Glu
130 135 140

Glu Arg Leu Arg Glu Met Phe Gly Leu Tyr Gly Gln Thr Thr Gly Lys
145 150 155 160

Gly Ser Ile Ser Leu Lys Glu Leu Asn Ala Arg Pro Leu Glu Val Phe
165 170 175

30 Met Cys Ser Val Leu Lys Arg Gln Gly Tyr Gly Glu Gly Phe Arg Trp
180 185 190

Met Ala Gln Tyr Ile Asp Xaa
195

(2) INFORMATION FOR SEQ ID NO: 449:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 258 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 449:

45 Met Thr Leu Ser Arg Phe Ala Tyr Asn Gly Lys Arg Cys Pro Ser Ser
1 5 10 15

Tyr Asn Ile Leu Asp Asn Ser Lys Ile Ile Ser Glu Glu Cys Arg Lys
20 25 30

Glu Leu Thr Ala Leu Leu His His Tyr Tyr Pro Ile Glu Ile Asp Pro
35 40 45

55 His Arg Thr Val Lys Glu Lys Leu Pro His Met Val Glu Trp Thr
50 55 60

Lys Ala His Asn Leu Cys Gln Lys Ile Gln Lys Phe Gln Ile
65 70 75 80

60 Ala Gln Val Val Arg Glu Ser Asn Ala Met Leu Arg Glu Gly Tyr Lys
85

622

5 Thr Phe Phe Asn Thr Leu Tyr His Asn Asn Ile Pro Leu Phe Ile Phe 85 90 95
100 105 110
Ser Ala Gly Ile Gly Asp Ile Leu Glu Ile Ile Arg Gln Met Lys 115 120 125
130 135 140
10 Val Phe His Pro Asn Ile His Ile Val Ser Asn Tyr Met Asp Phe Asn 145 150 155 160
Glu Asp Gly Phe Leu Gln Gly Phe Lys Gly Gln Leu Ile His Thr Tyr 165 170 175
15 Asn Lys Asn Ser Ser Val Cys Glu Asn Xaa Gly Tyr Phe Gln Gln Leu 180 185 190 195
Glu Gly Lys Thr Asn Val Ile Leu Leu Gly Asp Ser Ile Gly Asp Leu 200 205 210 215
20 Thr Met Ala Asp Gly Val Pro Gly Val Gln Asn Ile Leu Lys Ile Gly 220 225 230 235
Phe Leu Asn Asp Lys Val Glu Glu Arg Xaa Arg Tyr Met Asp Ser 240 245 250 255
25 Tyr Asp Ile Val Leu Glu Lys Asp Glu Thr Leu Asp Val Val Asn Gly 260 265 270 275
30 Leu Leu Gln His Ile Leu Cys Gln Gly Val Gln Leu Glu Met Gln Gly 280 285 290 295
Pro Xaa 300 305 310 315
35 (2) INFORMATION FOR SEQ ID NO: 450:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 450:
Met Ser His Val Leu Leu Cys Pro Ser Leu Ser Cys Ser Asn Leu Leu 1 5 10 15
Pro Pro Ser His Ser Leu Glu Thr Met Gly Ser Leu Ser Pro His Leu 20 25 30
Cys Gly His Thr Met Cys Pro Val Asn Pro Glu Leu Pro Leu Ser Ser 35 40 45
Arg Leu Thr Thr Asp Gln Pro Gln Pro Asp Ala Cys Ser Pro Thr Leu 50 55 60
Leu Thr Leu Pro Leu Pro Ser Ser Phe Leu Pro His Ser Lys Pro Thr 65 70 75 80

623

Phe Xaa His Pro Cys Ser Pro 85
5 (2) INFORMATION FOR SEQ ID NO: 451:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 451:
Met Phe Ser Ile Asn Pro Leu Glu Asn Leu Lys Val Tyr Ile Ser Ser 1 5 10 15
Arg Pro Leu Val Val Phe Met Ile Ser Val Xaa Pro Met Ala Ile 20 25 30
Ala Phe Leu Thr Leu Gly Tyr Phe Phe Lys Ile Lys Glu Ile Lys Ser 35 40 45
Pro Glu Met Ala Glu Asp Trp Asn Thr Phe Leu Leu Arg Phe Asn Asp 50 55 60
Leu Asp Leu Cys Val Ser Glu Asn Glu Thr Leu Lys His Leu Thr Asn 65 70 75 80
Asp Thr Thr Thr Pro Glu Ser Thr Met Thr Ser Gly Gln Ala Arg Ala 85 90 95
Ser Thr Gln Ser Pro Gln Ala Leu Glu Asp Ser Gly Pro Val Asn Ile 100 105 110 115
Ser Val Ser Ile Thr Leu Thr Leu Asp Pro Leu Lys Pro Phe Gly Gly 120 125 130 135
Tyr Ser Arg Asn Val Thr His Leu Tyr Ser Thr Ile Leu Gly His Gln 140 145 150 155
Ile Gly Leu Ser Gly Arg Glu Ala His Glu Glu Ile Asn Ile Thr Phe 160 165 170 175
Thr Leu Pro Thr Ala Trp Ser Ser Asp Asp Cys Ala Leu His Gly His 180 185 190 195
Cys Glu Gln Val Val Phe Thr Ala Cys Met Thr Leu Thr Ala Ser Pro 200 205 210 215
Gly Val Phe Pro Val Thr Val Gln Pro Pro His Cys Val Pro Asp Thr 220 225 230 235
Tyr Ser Asn Ala Thr Leu Trp Tyr Lys Ile Phe Thr Thr Ala Arg Asp 240 245 250 255
Ala Asn Thr Lys Tyr Ala Gln Asp Tyr Asn Pro Phe Trp Cys Tyr Lys 260 265 270 275
Gly Ala Ile Gly Lys Val Tyr His Ala Leu Asn Pro Lys Leu Thr Val 280 285 290 295

Ile Val Pro Asp Asp Arg Ser Leu Ile Asn Leu His Leu Met His
260 265 270
5 Thr Ser Tyr Phe Leu Phe Val Met Val Ile Thr Met Phe Cys Tyr Ala
275 280 285

Val Ile Lys Gly Arg Pro Ser Lys Leu Arg Gln Ser Asn Pro Glu Phe
290 295 300

Cys Pro Glu Lys Val Ala Leu Ala Glu Ala Xaa
305 310 315

(2) INFORMATION FOR SEQ ID NO: 452:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 452:

Met Pro Gly Leu Ser Leu Ala Leu Leu Pro Phe Gly Pro Gly Cys Thr
1 5 10 15

Glu Ala Leu His Ala Gly Cys Phe Pro Ala Phe Ala Ser Ala Thr Arg
20 25 30

Val Asn Gly Glu Ala Ala Leu Ser Pro Gly Leu Cys Asp Pro Ile Ser
35 40 45

Val Pro Tyr Val
50

35

(2) INFORMATION FOR SEQ ID NO: 453:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 383 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 453:

Met Ala Val Gly Gln Ile Met Thr Phe Gly Ser Pro Val Ile Gly Cys
1 5 10 15

Gly Phe Ile Ser Gly Trp Asn Leu Val Ser Met Cys Val Gly Tyr Val
20 25 30

Leu Leu Trp Lys Val Tyr Gln Lys Thr Pro Ala Leu Ala Val Lys Ala
35 40 45

Gly Leu Lys Glu Glu Thr Glu Leu Lys Gln Leu Asn Leu His Lys
50 55 60

Asp Thr Glu Pro Lys Pro Leu Glu Gly Thr His Leu Met Gly Val Lys
65 70 75 80

60

Asp Ser Asn Ile His Glu Leu Glu His Glu Gln Glu Pro Thr Cys Ala
85 90 95

Ser Gln Met Ala Glu Pro Phe Arg Thr Phe Arg Asp Gly Trp Val Ser
100 105 110

Tyr Tyr Asn Gln Pro Val Phe Leu Ala Gly Met Gly Leu Ala Phe Leu
115 120 125

Tyr Met Thr Val Leu Gly Phe Asp Cys Ile Thr Thr Gly Tyr Ala Tyr
130 135 140

Thr Gln Gly Leu Ser Gly Phe His Pro Gln Tyr Phe Asp Gly Ser Ile
145 150 155 160

Ser Tyr Asn Trp Asn Asn Gly Asn Cys Ser Phe Tyr Leu Ala Thr Ser
165 170 175

Lys Met Trp Phe Gly Ser Ala Gly Leu Ile Ser Gly Leu Ala Gln Leu
180 185 190

Ser Cys Leu Ile Leu Cys Val Ile Ser Val Phe Met Pro Gly Ser Pro
195 200 205

Leu Asp Leu Ser Val Ser Pro Phe Glu Asp Ile Arg Ser Arg Phe Ile
210 215 220

Gln Gly Glu Ser Ile Thr Pro Thr Lys Ile Pro Glu Ile Thr Thr Glu
225 230 235 240

Ile Tyr Met Ser Asn Gly Ser Asn Ser Ala Asn Ile Val Pro Glu Thr
245 250 255

Ser Pro Glu Ser Val Pro Ile Ile Ser Val Ser Leu Leu Phe Ala Gly
260 265 270

Val Ile Ala Ala Arg Ile Gly Leu Trp Ser Phe Asp Leu Thr Val Thr
275 280 285

Gln Leu Leu Gln Glu Asn Val Ile Glu Ser Glu Arg Gly Ile Ile Asn
290 295 300

Gly Val Gln Asn Ser Met Asn Tyr Leu Leu Asp Leu Leu His Phe Ile
305 310 315 320

Met Val Ile Leu Ala Pro Asn Pro Glu Ala Phe Gly Leu Leu Val Leu
325 330 335

Ile Ser Val Ser Phe Val Ala Met Gly His Ile Met Tyr Phe Arg Phe
340 345 350

Ala Gln Asn Thr Leu Gly Asn Lys Leu Phe Ala Cys Gly Pro Asp Ala
355 360 365

Lys Glu Val Arg Lys Glu Asn Gln Ala Asn Thr Ser Val Val Xaa
370 375 380

(2) INFORMATION FOR SEQ ID NO: 454:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 186 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 454:
- Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe 1 5 10 15
 Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met 20 25 30
 Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ile Thr Phe Pro 35 40 45
 Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly 50 55 60
 Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu 65 70 75 80
 Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu 85 90 95
 Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro 100 105 110 115
 Gln His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Phe Leu 120 125
 Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile 130 135 140 145
 Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Trp Arg Lys His 150 155 160
 Cys Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala 165 170 175
 Lys Glu Pro Leu Leu Gln Asp Thr Asn Val 180 185
- (2) INFORMATION FOR SEQ ID NO: 455:
- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 163 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 455:
- Met Leu Gln Thr Ser Asn Tyr Ser Leu Val Leu Ser Leu Gln Phe Leu 1 5 10 15
 Leu Leu Ser Tyr Asp Leu Phe Val Asn Ser Phe Ser Glu Leu Leu Gln 20 25 30
 Lys Thr Pro Val Ile Gln Leu Val Leu Phe Ile Ile Gln Asp Ile Ala

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 163 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 456:
- Met Arg Ile Gln Val Phe Ile Leu Leu Leu Gly Ala Gly Gly Thr Ser 1 5 10 15
 Gln Phe Thr Lys Pro Pro Ser Leu Pro Leu Glu Pro Glu Pro Ala Val 20 25 30
 Glu Ser Ser Pro Thr Glu Thr Ser Glu Gln Ile Arg Glu Lys 35 40 45
- (2) INFORMATION FOR SEQ ID NO: 457:
- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 105 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 457:
- Met Ser Tyr Leu Ala Phe Leu Thr Met Thr Phe Asp Phe Cys Cys Leu 1 5 10 15
 Tyr Phe Ser Thr Val Tyr Ala Pro Ser Phe Lys Tyr Ile Cys Val His 20 25 30
- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 163 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 456:
- Val Leu Phe Asn Ile Ile Ile Phe Leu Met Phe Phe Asn Thr Phe 35 40 45
 Val Phe Gln Ala Gly Leu Val Asn Leu Leu Phe His Lys Phe Lys Gly 50 55 60
 Thr Ile Ile Leu Thr Ala Val Tyr Phe Ala Leu Ser Ile Ser Leu His 65 70 75 80
 Val Trp Val Met Asn Leu Arg Trp Lys Asn Ser Asn Ser Phe Ile Trp 85 90 95 100 105 110 115
 Thr Asp Gly Leu Gln Met Leu Phe Val Phe Gln Arg Leu Ala Val 120 125
 Leu Tyr Cys Tyr Phe Tyr Lys Arg Thr Ala Val Arg Leu Gly Asp Pro 130 135 140 145
 His Phe Tyr Gln Asp Ser Leu Trp Leu Arg Lys Glu Phe Met Gln Val 150 155 160
 Arg Arg Xaa

Thr Asp Thr His Ile Cys Val Cys Val Cys Ile Tyr Leu Ser Ser Val
 35 45
 5 Val Ser Lys Ser Ser Ala Glu Ala Asp Gly Val Leu Gln Pro Arg Arg
 50 55 60
 His Pro Ala Ser Leu Leu Ile Val Phe Ala Thr Ser Ile Ser Glu Ser
 65 70 75 80
 10 Ser Leu Leu Ile Phe Ser Phe Gln Lys Thr Glu Ala Lys Leu Ile Val
 85 90 95
 Phe Ala Val Ser Leu Leu Ala Lys Xaa
 100 105

(2) INFORMATION FOR SEQ ID NO: 458:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 458:

Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser
 1 5 10 15
 30 Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val
 20 25 30
 Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro
 35 40 45
 Ser Ser Glu Leu Leu Glu Gln Leu Leu Ser Val Gln Phe Val Trp Trp Gln
 50 55 60
 Ala His Thr Val Ala Xaa
 65 70

(2) INFORMATION FOR SEQ ID NO: 459:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 155 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 459:

Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe
 1 5 10 15
 55 Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val
 20 25 30
 Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro
 35 40 45

Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala
 50 55 60
 5 Val Gly Phe Leu Glu Leu Ala Gly Leu Leu Val Met Gly Pro
 65 70 75 80
 Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met
 85 90 95
 10 Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu Ser Thr Cys
 100 105 110
 Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Asn Val Gly
 115 120 125
 15 Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys
 130 135 140
 Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa
 145 150 155

(2) INFORMATION FOR SEQ ID NO: 460:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 460:

Met Lys Leu Gly Arg Ala Val Leu Gly Leu Leu Leu Ala Pro Ser
 1 5 10 15
 35 Val Val Gln Ala Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly
 20 25 30
 Val Leu Thr Gly Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu
 35 40 45
 Cys Cys Gly Gln Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp
 50 55 60
 45 Leu Asp Asp Asn Leu Phe Gly Gln His Leu Ala Lys Ile Ile Leu
 65 70 75 80
 Asn Ala Val Phe Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu
 85 90 95
 50 Thr Leu Ser Leu His Gly Trp Thr Gly Thr Gly Lys Asn Phe Val Ser
 100 105 110
 Lys Ile Ile Ala Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr
 115 120 125
 55 Val His Leu Phe Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile
 130 135 140
 Thr Leu Tyr Lys Asp Gln Leu Leu Trp Ile Arg Gly Asn Val Ser
 145 150 155 160

Ala Cys Ala Arg Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His
165 170 175

5 Ala Gly Leu Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu
180 185 190

Val Asp Gly Val Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn
195 200 205

10 Ala Gly Ala Glu Arg Ile Thr Asp Val Ala Leu Asp Phe Tyr Arg Ser
210 215 220

Gly Lys Gln Arg Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu
225 230 235 240

Ser Val Ser Val Phe Asn Asn Lys Asn Ser Gly Phe Tyr His Ser Ser
245 250 255

20 Leu Ile Asp Arg Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu
260 265 270

Glu Tyr Lys His Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg
275 280 285

25 Gly Tyr Glu Ile Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met
290 295 300

Thr Phe Phe Pro Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys
305 310 315 320

Thr Val Phe Thr Lys Leu Asp Tyr Tyr Asp Asp
325 330

35

(2) INFORMATION FOR SEQ ID NO: 461:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 461:

40 Gly Lys Lys Cys Ile 5

45 Met Leu Lys Cys Ile 5

50

(2) INFORMATION FOR SEQ ID NO: 462:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 462:

55 Met Ile Leu Thr Leu Leu Ser Val Val Ser Thr Met Ala Ser
1 5 10

60

(2) INFORMATION FOR SEQ ID NO: 463:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 285 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 463:

5 Met Lys Leu His Pro Pro Pro Ser Pro Val Thr Gln Asp His Arg
1 5 10 15

10 Ser Lys Ser Ser His Ser Asn Trp Met Pro Arg Met Gly Ala Cys Ser
20 25 30

15 Met Ser Arg Thr Ser Ser Gly Pro Pro Ser Leu Cys Lys Ser Thr
35 40 45

20 Ser Gly Arg Ser Cys Thr Arg Pro His Cys Trp Pro Ser Leu Pro Ala
50 55 60

Trp Val Ser Val Phe Thr Arg Thr Asn Thr Gly Ser Trp Cys Tyr Pro
65 70 75 80

25 Ala Trp Gly Gly Ala Phe Ser Arg Pro Trp Met Ser Ala Gln Ser Met
85 90 95

30 Cys Cys Ala Glu Arg Ser Val Leu Gln Val Ala Cys Arg Leu Leu Asp
100 105 110

Ala Leu Glu Phe Leu His Glu Asn Glu Tyr Val His Gly Asn Val Thr
115 120 125

35 Ala Glu Asn Ile Phe Val Asp Pro Glu Asp Gln Ser Gln Val Thr Leu
130 135 140

Ala Gly Tyr Gly Phe Ala Phe Arg Tyr Cys Pro Ser Gly Lys His Val
145 150 155 160

40 Ala Tyr Val Glu Gly Ser Arg Ser Pro His Glu Gly Asp Leu Glu Phe
165 170 175

45 Ile Ser Met Asp Leu His Lys Gly Cys Gly Pro Ser Arg Arg Xaa Asp
180 185 190

Leu Gln Ser Leu Gly Tyr Cys Met Leu Lys Trp Leu Tyr Gly Phe Leu
195 200 205

50 Pro Trp Thr Asn Cys Leu Pro Xaa Xaa Glu Asp Ile Met Lys Gln Lys
210 215 220

Gln Lys Phe Val Asp Lys Pro Gly Pro Phe Val Gly Pro Cys Gly His
225 230 235 240

55 Trp Ile Arg Pro Ser Glu Thr Leu Gln Lys Tyr Leu Lys Val Val Met
245 250 255

60 Ala Leu Thr Tyr Glu Glu Lys Pro Pro Tyr Ala Met Leu Arg Asn
260 265 270

632

633

Leu Glu Ala Leu Leu Gln Asp Leu Arg Val Ser Pro Tyr
275 280 285

5

(2) INFORMATION FOR SEQ ID NO: 464:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 464:

Met Thr Ser Pro Pro His Gln Gly Trp Glu Gln Arg Gly Cys Gly
1 5 10 15

Glu Ser Gln Val Pro Leu Ala Leu Ser Arg Val Phe Ser Thr Ser His
20 25 30

Tyr Cys Leu Leu Val Ala Asn Gln Ser Ile Phe Phe Pro Cys Leu
35 40 45

Trp Ala Val Glu Arg Leu Leu Gly Val Arg Cys Thr Cys Pro Leu Ser
50 55 60

Trp Gly Lys Arg Ile Ile Ser Glu His Cys Ser Ala Gln Ser Ser Xaa
65 70 75 80

30

(2) INFORMATION FOR SEQ ID NO: 465:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 465:

Met His Thr Trp Tyr Asn Asp Arg Arg Gln Asn Cys His Cys Leu Leu
1 5 10 15

Phe Phe Leu Ile Tyr Leu Arg Lys Ile Tyr Gln Val Val Pro His Val
20 25 30

Pro Leu Leu Val Lys Cys Arg Gly Arg Leu Lys Gly Val Asn Ile
35 40 45

(2) INFORMATION FOR SEQ ID NO: 466:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 466:

Met Glu Leu Val Leu Val Phe Leu Cys Ser Leu Leu Ala Pro Met Val
1 5 10 15

Leu Ala Ser Ala Ala Glu Lys Glu Lys Glu Met Asp Pro Phe His Tyr
20 25 30

Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu Val Phe Ala Val Val Leu
35 40 45

Phe Ser Val Gly Ile Leu Leu Ile Leu Ser Arg Arg Cys Lys Cys Ser
50 55 60

Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp Glu Glu Ala Gln Val Glu
65 70 75 80

Asn Leu Ile Thr Ala Asn Ala Thr Glu Glu Pro Gln Lys Ala Glu Asn Xaa
85 90 95

20

(2) INFORMATION FOR SEQ ID NO: 467:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 399 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 467:

Met Ala Ser Gly Ala Asp Ser Lys Gly Asp Asp Leu Ser Thr Ala Ile
1 5 10 15

Leu Lys Gln Lys Asn Arg Pro Asn Arg Leu Ile Val Asp Glu Ala Ile
20 25 30

Asn Glu Asp Asn Ser Val Val Ser Leu Ser Gln Pro Lys Met Asp Glu
35 40 45

Leu Gln Leu Phe Arg Gly Asp Thr Val Leu Leu Lys Gly Lys Lys Arg
50 55 60

Arg Glu Ala Val Cys Ile Val Leu Ser Asp Asp Thr Cys Ser Asp Glu
65 70 75 80

Lys Ile Arg Met Asn Arg Val Val Arg Asn Asn Leu Arg Val Arg Leu
85 90 95

Gly Asp Val Ile Ser Ile Gln Pro Cys Pro Asp Val Lys Tyr Gly Lys
100 105 110

Arg Ile His Val Leu Pro Ile Asp Asp Thr Val Glu Gly Ile Thr Gly
115 120 125

Asn Leu Phe Glu Val Tyr Leu Lys Pro Tyr Phe Leu Glu Ala Tyr Arg
130 135 140

Pro Ile Arg Lys Gly Asp Ile Phe Leu Val Arg Gly Gly Met Arg Ala
145 150 155

145 150 155 160
Val Glu Phe Lys Val Val Glu Thr Asp Pro Ser Pro Tyr Cys Ile Val
165 170 175

5

Ala Pro Asp Thr Val Ile His Cys Glu Glu Pro Ile Lys Arg Glu
180 185 190

Asp Glu Glu Glu Ser Leu Asn Glu Val Gly Tyr Asp Asp Ile Gly Gly
195 200 205

10

Cys Arg Lys Glu Leu Ala Gln Ile Lys Glu Met Val Glu Leu Pro Leu
210 215 220

15

Arg His Pro Ala Leu Phe Lys Ala Ile Gly Val Lys Pro Arg Gly
225 230 235 240

Ile Leu Leu Tyr Gly Pro Pro Gly Thr Gly Lys Thr Leu Ile Ala Arg
245 250 255

20

Ala Val Ala Asn Glu Thr Gly Ala Phe Phe Leu Ile Asn Gly Pro
260 265 270

25

Glu Ile Met Ser Lys Leu Ala Gly Glu Ser Glu Ser Asn Leu Arg Lys
275 280 285

Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro Ala Ile Ile Phe Ile Asp
290 295 300

30

Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu Lys Thr His Gly Glu Val
305 310 315 320

Glu Arg Arg Ile Val Ser Gln Leu Leu Thr Met Asp Gly Leu Lys
325 330 335

35

Gln Arg Ala His Val Ile Val Met Ala Ala Thr Asn Arg Pro Asn Ser
340 345 350

40

Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg Phe Asp Arg Glu Val Asp
355 360 365

Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu Glu Ile Leu Gln Ile His
370 375 380

45

Thr Lys Asn Met Lys Leu Ala Asp Asp Val Asp Leu Glu Gln Xaa
385 390 395

(2) INFORMATION FOR SEQ ID NO: 468:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 468:

Leu
1

60

(2) INFORMATION FOR SEQ ID NO: 469:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 273 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 469:

Met Ala Ala Pro Lys Gly Ser Leu Trp Val Arg Thr Gln Leu Gly Leu
1 5 10 15

Pro Pro Leu Leu Leu Thr Met Ala Leu Ala Gly Gly Ser Gly Thr
20 25 30

Ala Ser Ala Glu Ala Phe Asp Ser Val Leu Gly Asp Thr Ala Ser Cys
35 40 45

His Arg Ala Cys Gln Leu Thr Tyr Pro Leu His Thr Tyr Pro Lys Glu
50 55 60

Glu Glu Leu Tyr Ala Cys Gln Arg Gly Cys Arg Leu Phe Ser Ile Cys
65 70 75 80

Gln Phe Val Asp Asp Gly Ile Asp Leu Asn Arg Thr Lys Leu Glu Cys
85 90 95

Glu Ser Ala Cys Thr Glu Ala Tyr Ser Gln Ser Asp Glu Gln Tyr Ala
100 105 110

Cys His Leu Gly Cys Gln Asn Gln Leu Pro Phe Ala Glu Leu Arg Gln
115 120 125

Glu Gln Leu Met Ser Leu Met Pro Lys Met His Leu Leu Phe Pro Leu
130 135 140

Thr Leu Val Arg Ser Phe Trp Ser Asp Met Met Asp Ser Ala Gln Ser
145 150 155 160

Phe Ile Thr Ser Ser Trp Thr Phe Tyr Leu Gln Ala Asp Asp Gly Lys
165 170 175

Ile Val Ile Phe Xaa Ser Lys Pro Arg Asn Pro Arg Tyr Ala Pro His
180 185 190

Leu Glu Pro Gly Ala Leu Leu Pro Asn Leu Xaa Xaa Ser Leu Ser Lys
195 200 205

Met Ser Xaa Xaa Ser Xaa Met Arg Asn Ser Gln Ala His Arg Asn Phe
210 215 220

Leu Glu Asp Gly Glu Ser Asp Gly Phe Leu Arg Cys Leu Ser Leu Asn
225 230 235 240

Ser Gly Trp Ile Leu Thr Thr Thr Leu Val Leu Ser Val Met Val Leu
245 250 255

Leu Trp Ile Cys Cys Ala Thr Cys Cys Tyr Thr Leu Leu Asp Ala Val
260 265 270

Xaa

5

(2) INFORMATION FOR SEQ ID NO: 470:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 470:

15 Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe Ser 15
1 5 10

Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr Pro 30
20 25

20 Phe Ile Gly Pro Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile Ala 45
35 40

25 Thr Glu Lys Arg Leu Thr Lys Leu Val His Ser Ser Leu Val Gly 60
50 55

Ser Ile Leu Ser Ala Leu Ser Ala Leu Val Gly Phe Ile Ile Leu Ser 80
65 70 75

30 Val Lys Gln Ala Thr Leu Asn Pro Ala Ser Leu Gln Cys Glu Leu Asp 95
85 90

Lys Asn Asn Ile Pro Thr Arg Ser Tyr Val Ser Tyr Phe Tyr His Asp 110
100 105

35 Ser Leu Tyr Thr Thr Asp Cys Tyr Thr Ala Lys Ala Ser Leu Ala Gly 125
115 120

40 Xaa Leu Ser Leu Met Leu Ile Cys Thr Leu Leu Glu Phe Cys Leu Ala 140
130 135

Val Leu Thr Ala Val Leu Arg Trp Lys Gln Ala Tyr Ser Asp Phe Pro 160
145 150 155

45 Gly Ser Val Leu Phe Leu Pro His Ser Tyr Tyr Ile Gly Asn Ser Gly Met 175
165 170

Ser Ser Lys Met Thr His Asp Cys Gly Tyr Glu Glu Leu Thr Ser 190
180 185

50

55

(2) INFORMATION FOR SEQ ID NO: 471:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 234 amino acids

(B) TYPE: amino acid

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 471:

5 Met Arg Lys Thr Arg Leu Trp Gly Leu Leu Trp Met Leu Phe Val Ser 15
1 5 10

Glu Leu Arg Ala Ala Thr Lys Leu Thr Glu Glu Lys Tyr Glu Leu Lys 30
20 25

10 Glu Gly Gln Thr Leu Asp Val Lys Cys Asp Tyr Thr Leu Glu Lys Phe 45
35 40 45

15 Ala Ser Ser Gln Lys Ala Trp Gln Ile Ile Arg Asp Gly Glu Met Pro 60
50 55 60

Lys Thr Leu Ala Cys Thr Glu Arg Pro Ser Lys Asn Ser His Pro Val 80
65 70 75

20 Gln Val Gly Arg Ile Ile Leu Glu Asp Tyr His Asp His Gly Leu Leu 95
85 90

Arg Val Arg Met Val Asn Leu Gln Val Glu Asp Ser Gly Leu Tyr Gln 110
100 105

25 Cys Val Ile Tyr Gln Pro Pro Lys Glu Pro His Met Leu Phe Asp Arg 125
115 120

30 Ile Arg Leu Val Val Thr Lys Gly Phe Ser Gly Thr Pro Gly Ser Asn 140
130 135 140

Glu Asn Ser Thr Gln Asn Val Tyr Lys Ile Pro Pro Thr Thr Thr Lys 160
145 150 155

35 Ala Leu Cys Pro Leu Tyr Thr Ser Pro Arg Thr Val Thr Gln Ala Pro 175
165 170

Pro Lys Ser Thr Ala Asp Val Ser Thr Pro Asp Ser Glu Ile Asn Leu 190
180 185

40 Thr Asn Val Thr Asp Ile Ile Arg Val Pro Val Phe Asn Ile Val Ile 205
195 200

Leu Leu Ala Gly Gly Phe Leu Ser Lys Ser Leu Val Phe Ser Val Leu 220
210 215 220

Phe Ala Val Thr Leu Arg Ser Phe Val Pro 230
225

50

(2) INFORMATION FOR SEQ ID NO: 472:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 105 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 472:

60 Met Leu His Ile Leu Pro Leu Lys Ser Tyr Asp Phe Pro His Phe Ser 15
1 5 10

Leu Met Gly Arg Tyr Arg Cys Ala Ser Leu Leu Phe Cys Phe Leu Leu
 20 25 30
 5 Leu Phe Phe Phe Cys Ser Val Leu Trp Thr Phe Ser Asp Met His
 35 40 45
 Arg Ser Gly Glu Asp Gly Pro Trp Thr Pro Cys Val His His Leu Ala
 50 55 60
 10 Ala Ser Leu Ile Ser Tyr Gly Cln Pro Gly Phe Ile Cys Ile Ser Leu
 65 70 75 80
 Phe Ser Pro Val Leu Phe Ile Glu Asn Pro Arg His Tyr Ala Asn Ala
 85 90 95
 Thr Val Thr Thr Leu Gly Asp Trp Xaa
 100 105

(2) INFORMATION FOR SEQ ID NO: 473:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 amino acids
 (B) TYPE: amino acid
 (C) TOPOLOGY: linear
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 473:

Met Val Phe Leu Lys Tyr Arg Phe Phe Leu Val Phe Leu Ala
 1 5 10 15
 30 Asn Cys Ile Tyr Ser Leu His Tyr Lys Pro Ser Leu Met Tyr Pro Lys
 20 25 30

(2) INFORMATION FOR SEQ ID NO: 474:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 571 amino acids
 (B) TYPE: amino acid
 (C) TOPOLOGY: linear
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 474:

Met Ala Leu Ser Arg Gly Leu Pro Arg Glu Leu Ala Glu Ala Val Ala
 1 5 10 15
 Gly Gly Arg Val Leu Val Val Gly Ala Gly Gly Ile Gly Cys Glu Leu
 20 25 30
 55 Leu Lys Asn Leu Val Leu Thr Gly Phe Ser His Ile Asp Leu Ile Asp
 35 40 45
 Leu Asp Thr Ile Asp Val Ser Asn Leu Asn Arg Cln Phe Leu Phe Cln
 50 55 60

Lys Lys His Val Gly Arg Ser Lys Ala Gln Val Ala Lys Glu Ser Val
 65 70 75 80
 5 Leu Cln Phe Tyr Pro Lys Ala Asn Ile Val Ala Tyr His Asp Ser Ile
 85 90 95
 Met Asn Pro Asp Tyr Asn Val Glu Phe Arg Gln Phe Ile Leu Val
 100 105 110
 10 Met Asn Ala Leu Asp Asn Arg Ala Ala Arg Asn His Val Asn Arg Met
 115 120 125
 Cys Leu Ala Ala Asp Val Pro Leu Ile Glu Ser Gly Thr Ala Gly Tyr
 130 135 140
 15 Leu Gly Cln Val Thr Thr Ile Lys Lys Gly Val Thr Glu Cys Tyr Glu
 145 150 155 160
 Cys His Pro Lys Pro Thr Gln Arg Thr Phe Pro Gly Cys Thr Ile Arg
 165 170 175
 20 Asn Thr Pro Ser Glu Pro Ile His Cys Ile Val Trp Ala Lys Tyr Leu
 180 185 190
 25 Phe Asn Cln Leu Phe Gly Glu Asp Ala Asp Cln Glu Val Ser Pro
 195 200 205
 Asp Arg Ala Asp Pro Glu Ala Ala Trp Glu Pro Thr Glu Ala Glu Ala
 210 215 220
 30 Arg Ala Arg Ala Ser Asn Glu Asp Gly Asp Ile Lys Arg Ile Ser Thr
 225 230 235
 35 Lys Glu Trp Ala Lys Ser Thr Gly Tyr Asp Pro Val Lys Leu Phe Thr
 240 245 250 255
 Lys Leu Phe Lys Asp Asp Ile Arg Tyr Leu Leu Thr Met Asp Lys Leu
 260 265 270
 40 Trp Arg Lys Arg Lys Pro Pro Val Pro Leu Asp Trp Ala Glu Val Cln
 275 280 285
 Ser Cln Gly Glu Thr Asn Ala Ser Asp Cln Gln Asn Glu Pro Cln
 290 295 300
 45 Leu Gly Leu Lys Asp Cln Gln Val Leu Asp Val Lys Ser Tyr Ala Arg
 305 310 315 320
 Leu Phe Ser Lys Ser Ile Glu Thr Leu Arg Val His Leu Ala Glu Lys
 325 330 335
 Gly Asp Gly Ala Glu Leu Ile Trp Asp Lys Asp Asp Pro Ser Ala Met
 340 345 350
 55 Asp Phe Val Thr Ser Ala Ala Asn Leu Arg Met His Ile Phe Ser Met
 355 360 365
 Asn Met Lys Ser Arg Phe Asp Ile Lys Ser Met Ala Gly Asn Ile Ile
 370 375 380
 60

Pro Ala Ile Ala Thr Thr Asn Ala Val Ile Ala Gly Leu Ile Val Val
385 390 395 400

Glu Gly Leu Lys Ile Leu Ser Gly Lys Ile Asn Gln Cys Arg Thr Ile
405 410 415

Phe Leu Asn Lys Gln Pro Asn Pro Arg Lys Lys Leu Leu Val Pro Cys
420 425 430

Ala Leu Asp Pro Pro Asn Pro Asn Cys Tyr Val Cys Ala Ser Lys Pro
435 440 445

Glu Val Thr Val Arg Leu Asn Val His Lys Val Thr Val Leu Thr Leu
450 455 460

Gln Asp Lys Ile Val Lys Glu Lys Phe Ala Met Val Ala Pro Asp Val
465 470 475 480

Gln Ile Glu Asp Gly Lys Gly Thr Ile Leu Ile Ser Ser Glu Glu Gly
485 490 495

Glu Thr Glu Ala Asn Asn His Lys Lys Leu Ser Glu Phe Gly Ile Arg
500 505 510

Asn Gly Ser Arg Leu Gln Ala Asp Asp Phe Leu Gln Asp Tyr Thr Leu
515 520 525

Leu Ile Asn Ile Leu His Ser Glu Asp Leu Gly Lys Asp Val Glu Phe
530 535 540

Glu Val Val Gly Asp Ala Pro Glu Lys Val Gly Xaa Lys Gln Ala Glu
545 550 555 560

Asp Ala Ala Lys Ser Ile Thr Asn Gly Gln Xaa
565 570

(2) INFORMATION FOR SEQ ID NO: 475:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 475:

Met Gln Val Val Thr Cys Leu Thr Arg Asp Ser Tyr Leu Thr His Cys
1 5 10 15

Phe Leu Gln His Leu Met Val Val Leu Ser Ser Leu Glu Arg Thr Pro
20 25 30

Ser Pro Glu Pro Val Asp Lys Asp Phe Tyr Ser Glu Phe Gly Asn Lys
35 40 45

Thr Thr Gly Lys Met Glu Asn Tyr Glu Leu Ile His Ser Ser Arg Val
50 55 60

Lys Phe Thr Tyr Pro Ser Glu Glu Ile Gly Asp Leu Thr Phe Thr
65 70 75 80

Val Ala Gln Lys Met Ala Glu Pro Glu Lys Ala Pro Ala Leu Ser Ile
85 90 95

Leu Leu Tyr Val Gln Ala Phe Gln Val Gly Met Pro Pro Gly Cys
100 105 110

Cys Arg Gly Pro Leu Arg Pro Lys Thr Leu Leu Thr Ser Ser Glu
115 120 125

Ile Phe Leu Leu Asp Glu Asp Cys Val His Tyr Pro Leu Pro Glu Phe
130 135 140

Ala Lys Glu Pro Pro Gln Arg Asp Tyr Arg Leu Asp Asp Gly Arg
145 150 155 160

Arg Val Arg Asp Leu Asp Arg Val Leu Met Gly Tyr Gln Thr Tyr Pro
165 170 175

Gln Pro Ser Pro Ser Ser Met Thr Cys Lys Val Met Thr Ser Trp
180 185 190

Ala Val Ser Pro Trp Thr Thr Leu Gly Arg Cys Gln Val Ala Arg Leu
195 200 205

Glu Pro Ala Arg Ala Val Lys Ser Ser Gly Arg Cys Leu Ser Pro Val
210 215 220

Leu Arg Ala Glu Arg Ser Ser Arg Cys Trp Leu Ala Ser Gly Arg
225 230 235 240

Pro Cys Val Ala Val Ser Cys Leu Ser Ser Pro Ala Ser Pro Gly
245 250 255

His Ser Gln Pro Val Val Ser Ser Leu Thr Pro Thr Gly Ala Gly Gln
260 265 270

Gln Ala Phe Val Phe Ser Lys Asn Val Leu Ser Ser Leu Trp Tyr Leu
275 280 285

Asn Leu Thr Val Leu Ala Glu Asn Val Asn Met Cys Val Cys Cys Val
290 295 300 305

Asn Ser Phe Ser Cys Trp Glu Xaa
310

(2) INFORMATION FOR SEQ ID NO: 476:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 329 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 476:

Met Ala Gln His His Leu Trp Ile Leu Leu Leu Cys Leu Gln Thr Trp
1 5 10 15

Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile
20 25 30 35 40 45 50 55 60 65

5 Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln 30
35 45
Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr 60
50 60
Pro Gly Asp Ser Glu Thr Ala Pro Val Val Thr Val Thr His Arg Asn 80
65 75
Tyr Tyr Glu Arg Ile His Ala Leu Gly Pro Asn Tyr Asn Leu Val Ile 95
85 90
Ser Asp Leu Arg Met Glu Asp Ala Gly Asp Tyr Lys Ala Asp Ile Asn 110
100 110
Thr Gln Ala Asp Pro Tyr Thr Thr Lys Arg Tyr Asn Leu Gln Ile 125
115 120
Tyr Arg Arg Leu Gly Lys Pro Lys Ile Thr Gln Ser Leu Met Ala Ser 140
130 140
Val Asn Ser Thr Cys Asn Val Thr Leu Thr Cys Ser Val Glu Lys Glu 160
145 155
Glu Lys Asn Val Thr Tyr Asn Trp Ser Pro Leu Gly Glu Gly Asn 175
165 170
Val Leu Gln Ile Phe Gln Thr Pro Glu Asp Gln Glu Leu Thr Tyr Thr 190
180 185
Cys Thr Ala Gln Asn Pro Val Ser Asn Asn Ser Asp Ser Ile Ser Ala 205
195 200
Arg Gln Leu Cys Ala Asp Ile Ala Met Gly Phe Arg Thr His His Thr 220
210 215
Gly Leu Leu Ser Val Leu Ala Met Phe Phe Leu Leu Val Leu Ile Leu 240
225 230 235
Ser Ser Val Phe Leu Phe Arg Leu Phe Lys Arg Arg Gln Asp Ala Ala 255
245 250
Ser Lys Lys Thr Ile Tyr Thr Tyr Ile Met Ala Ser Arg Asn Thr Gln 270
260 265
Pro Ala Glu Ser Arg Ile Tyr Asp Glu Ile Leu Gln Ser Lys Val Leu 285
275 280
Pro Ser Lys Glu Glu Pro Val Asn Thr Val Tyr Ser Glu Val Gln Phe 300
290 295
Ala Asp Lys Met Gly Lys Ala Ser Thr Gln Asp Ser Lys Pro Pro Gly 320
305 310 315
Thr Ser Ser Tyr Glu Ile Val Ile Xaa 325

(2) INFORMATION FOR SEQ ID NO: 477:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 178 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 477:

5 Met Lys Leu Gln Cys Val Ser Leu Trp Leu Leu Gly Thr Ile Leu Ile 15
1 5 10
Leu Cys Ser Val Asp Asn His Gly Leu Arg Arg Cys Leu Ile Ser Thr 30
20 25
Asp Met His His Ile Glu Glu Ser Phe Gln Glu Ile Lys Arg Ala Ile 45
35 40
Gln Ala Lys Asp Thr Phe Pro Asn Val Thr Ile Leu Ser Thr Leu Glu 60
50 55
Thr Leu Gln Ile Ile Lys Pro Leu Asp Val Cys Cys Val Thr Lys Asn 80
65 70 75
Leu Leu Ala Phe Tyr Val Asp Arg Val Phe Lys Asp His Gln Glu Pro 95
85 90
Asn Pro Lys Ile Leu Arg Lys Ile Ser Ser Ile Ala Asn Ser Phe Leu 110
100 105
Tyr Met Gln Lys Thr Leu Arg Gln Cys Gln Glu Gln Arg Cys His 125
115 120
Cys Arg Gln Glu Ala Thr Asn Ala Thr Arg Val Ile His Asp Asn Tyr 140
130 135
Asp Gln Leu Glu Val His Ala Ala Ile Lys Ser Leu Gly Glu Leu 160
145 150 155
Asp Val Phe Leu Ala Trp Ile Asn Lys Asn His Glu Val Met Ser Ser 175
165 170
Ala Xaa

(2) INFORMATION FOR SEQ ID NO: 478:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 478:

55 Asp Thr Ala Ile Arg Val Ala Leu Ala Val Ala Val Leu Lys Thr Val 15
1 5 10
Ile Leu Gly Leu Leu Cys Leu Leu Cys Gly Gly Glu Gly Lys 30
20 25

Val Ala Gly Arg Gln Ala Val Thr Ser Asp Gln Gln Ser Val Gly Arg
35 40 45

5 Arg Asp Val Tyr
50

10 (2) INFORMATION FOR SEQ ID NO: 479:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 62 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 479:

Met Gln Lys Lys Asn Ser Leu Phe Phe Phe Ala Phe Tyr Tyr Glu
1 5 10 15

20 Asn Lys Thr Asn Ala Pro Gly Glu Gly Ser Met Ile Thr Arg Asn Ile
20 25 30

25 Lys Glu Tyr Phe Leu Pro Phe Leu Phe Cys Cys Val Glu Ala Ser Ile
35 40 45

Ala Ile Asn Lys Leu Asn Tyr Leu His Trp Thr His Phe Gln
50 60

30

(2) INFORMATION FOR SEQ ID NO: 480:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

40 Met Pro Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser
1 5 10 15

Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly
20 25

45

(2) INFORMATION FOR SEQ ID NO: 481:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 339 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

Met Ser Gly Pro Asp Val Glu Thr Pro Ser Ala Ile Gln Ile Cys Arg
1 5 10 15

Ile Met Arg Pro Asp Asp Ala Asn Val Ala Gly Asn Val His Gly Gly
20 25 30

60

Thr Ile Leu Lys Met Ile Glu Glu Ala Gly Ala Ile Ile Ser Thr Arg
35 40 45

5 His Cys Asn Ser Gln Asn Gly Glu Arg Cys Val Ala Ala Leu Ala Arg
50 55 60

10 Val Glu Arg Thr Asp Phe Leu Ser Pro Met Cys Ile Gly Glu Val Ala
65 70 75 80

His Val Ser Ala Glu Ile Thr Tyr Thr Ser Lys His Ser Val Glu Val
85 90 95

15 Gln Val Asn Val Met Ser Glu Asn Ile Leu Thr Gly Ala Lys Lys Leu
100 105 110

Thr Asn Lys Ala Thr Leu Trp Tyr Val Pro Leu Ser Leu Lys Asn Val
115 120 125

20 Asp Lys Val Leu Glu Val Pro Pro Val Val Tyr Ser Arg Xaa Glu Gln
130 135 140

Glu Glu Glu Gly Arg Lys Arg Tyr Glu Ala Gln Lys Leu Glu Arg Met
145 150 155 160

25 Glu Thr Lys Trp Arg Asn Gly Asp Ile Val Gln Pro Val Leu Asn Pro
165 170 175

30 Glu Pro Asn Thr Val Ser Tyr Ser Gln Ser Ser Leu Ile His Leu Val
180 185 190

Gly Pro Ser Asp Cys Thr Leu His Gly Phe Val His Gly Gly Val Thr
195 200 205

35 Met Lys Leu Met Asp Glu Val Ala Gly Ile Val Ala Ala Arg His Cys
210 215 220

Lys Thr Asn Ile Val Thr Ala Ser Val Asp Ala Ile Asn Phe His Asp
225 230 235 240

40 Lys Ile Arg Lys Gly Cys Val Ile Thr Ile Ser Gly Arg Met Thr Phe
245 250 255

Thr Ser Asn Lys Ser Met Glu Ile Glu Val Leu Val Asp Ala Asp Pro
260 265 270

Val Val Asp Ser Ser Gln Lys Arg Tyr Arg Ala Ala Ser Ala Phe Phe
275 280 285

50 Thr Tyr Val Ser Leu Ser Gln Glu Gly Arg Ser Leu Pro Val Pro Gln
290 295 300

Leu Val Pro Glu Thr Glu Asp Glu Lys Arg Phe Glu Glu Gly Lys
305 310 315 320

55 Gly Arg Tyr Leu Gln Met Lys Ala Lys Xaa Gln Gly His Ala Xaa Xaa
325 330 335

Gln Pro Xaa

60

646

647

(2) INFORMATION FOR SEQ ID NO: 482:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

Met Leu Asn Ser Asn Ile Asn Asp Leu Leu Met Val Thr Tyr Leu Ala
1 5 10 15
Asn Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn Leu
20 25 30

(2) INFORMATION FOR SEQ ID NO: 483:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 483:

Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu
1 5 10 15
Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr Arg Ile Gly
20 25 30
Cys Phe Lys Thr Ile Thr Cys Trp Pro Thr Ser Leu Thr Gln Arg Xaa
35 40 45

(2) INFORMATION FOR SEQ ID NO: 484:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 484:

Met Tyr Met Tyr Ser Leu Asn Val Phe Leu Ser Phe Ile Phe Leu Ala
1 5 10 15
Leu Val Phe Lys Cys Val His Val Cys Gln Gly Ala Asn Ala Phe Leu
20 25 30
Phe Leu Lys Leu Val Phe
35

(2) INFORMATION FOR SEQ ID NO: 485:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 485:

Met Gly Leu Arg Leu Ile Cys Leu Glu Leu Thr Met Val Lys Ala Leu
1 5 10 15
Val Cys Glu Met Phe Leu Phe Leu Met Thr Gln Lys Leu Ile Trp
20 25 30
Gln Glu Cys Thr Glu Lys Phe Ala Lys Leu Leu Val Gln Leu Ile Ser
35 40 45
Leu Val Phe Ala Trp Glu Phe Ser Glu Asp Thr Pro
50 55 60

(2) INFORMATION FOR SEQ ID NO: 486:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 486:

Met Leu Ala Ala Arg Leu Val Cys Leu Arg Thr Leu Pro Ser Arg Val
1 5 10 15
Phe His Pro Ala Phe Thr Lys Ala Ser Pro Val Val Lys Asn Ser Ile
20 25 30
Thr Lys Asn Gln Trp Leu Leu Thr Pro Ser Arg Glu Tyr Ala Thr Lys
35 40 45
Thr Arg Ile Gly Ile Arg Arg Gly Arg Thr Gly Gln Glu Leu Lys Glu
50 55 60
Ala Ala Leu Glu Pro Ser Met Glu Lys Ile Phe Lys Ile Asp Gln Met
65 70 75 80
Gly Arg Trp Phe Val Ala Gly Gly Ala Ala Val Gly Leu Gly Ala Leu
85 90 95
Cys Tyr Tyr Gly Leu Gly Leu Ser Asn Glu Ile Gly Ala Ile Glu Lys
100 105 110 115
Ala Val Ile Trp Pro Gln Tyr Val Lys Asp Arg Ile His Ser Thr Tyr
120 125 130
Met Tyr Leu Ala Gly Ser Ile Gly Leu Thr Ala Leu Ser Ala Ile Ala
135 140 145

Ile Ser Arg Thr Pro Val Leu Met Asn Phe Met Met Arg Gly Ser Trp
145 150 155 160

Val Thr Ile Gly Val Thr Phe Ala Ala Met Val Gly Ala Gly Met Leu
165 170 175

Val Arg Ser Ile Pro Tyr Asp Gln Ser Pro Gly Pro Lys His Leu Ala
180 185 190

10 Trp Leu Leu His Ser Gly Val Met Gly Ala Val Ala Pro Leu Thr
195 200 205

Ile Leu Gly Gly Pro Leu Leu Ile Arg Ala Ala Trp Tyr Thr Ala Gly
210 215 220

15 Ile Val Gly Gly Leu Ser Thr Val Ala Met Cys Ala Pro Ser Gly Lys
225 230 235 240

Phe Leu Asn Met Gly Ala Pro Leu Gly Val Gly Leu Gly Leu Val Phe
245 250 255

20 Val Ser Ser Leu Gly Ser Met Phe Leu Pro Pro Thr Thr Val Ala Gly
260 265 270

25 Ala Thr Tyr Ser Val Ala Met Tyr Gly Gly Leu Val Leu Phe Ser
275 280 285

Met Phe Leu Leu Tyr Asp Thr Gln Lys Val Ile Lys Arg Ala Gly Val
290 295 300

30 Ser Pro Met Tyr Gly Val Gln Lys Tyr Asp Pro Ile Asn Ser Met Leu
305 310 315 320

35 Ser Ile Tyr Met Asp Thr Leu Asn Ile Phe Met Arg Val Ala Thr Met
325 330 335

Leu Ala Thr Gly Gly Asn Arg Lys Lys Xaa.
340 345

40

(2) INFORMATION FOR SEQ ID NO: 487:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 487:

50 Met Glu Glu Val Leu Leu Leu Gly Leu Lys Asp Arg Gly Gly Tyr Thr
1 5 10 15

Ser Phe Trp Asn Asp Cys Ile Ser Ser Gly Leu Arg Gly Cys Met Leu
20 25 30

55 Ile Glu Leu Ala Leu Arg Gly Arg Leu Gln Leu Glu Ala Cys Gly Met
35 40 45

60 Arg Arg Lys Ser Leu Leu Thr Arg Lys Val Ile Cys Lys Ser Asp Ala
50 55 60

Pro Thr Gly Asp Val Leu Leu Asp Glu Ala Leu Lys His Val Lys Glu
65 70 75 80

5 Thr Gln Pro Pro Glu Thr Val Gln Asn Trp Ile Glu Leu Leu Ser Gly
85 90 95

Glu Thr Trp Asn Pro Leu Lys Leu His Tyr Gln Leu Arg Asn Val Arg
100 105 110

10 Glu Arg Leu Ala Lys Asn Leu Val Glu Lys Gly Val Leu Thr Thr Glu
115 120 125

15 Lys Gln Asn Phe Leu Phe Asp Met Thr Thr His Pro Leu Thr Asn
130 135 140

Asn Asn Ile Lys Gln Arg Leu Ile Lys Lys Val Gln Glu Ala Val Leu
145 150 155 160

20 Asp Lys Trp Val Asn Asp Pro His Arg Met Asp Arg Arg Leu Leu Ala
165 170 175

25 Leu Ile Tyr Leu Ala His Ala Ser Asp Val Leu Glu Asn Ala Phe Ala
180 185 190

Pro Leu Leu Asp Glu Gln Tyr Asp Leu Ala Thr Lys Arg Val Arg Gln
195 200 205

30 Leu Leu Asp Leu Asp Pro Glu Val Glu Cys Leu Lys Ala Asn Thr Asn
210 215 220

Glu Val Leu Trp Ala Val Val Ala Phe Thr Lys Xaa
225 230 235

35

(2) INFORMATION FOR SEQ ID NO: 488:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 200 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 488:

45 Met Ala Gln Arg Met Val Trp Val Asp Leu Glu Met Thr Gly Leu Asp
1 5 10 15

Ile Glu Lys Asp Gln Ile Ile Glu Met Ala Cys Leu Ile Thr Asp Ser
20 25 30

50 Asp Leu Asn Ile Leu Ala Glu Gly Pro Asn Leu Ile Ile Lys Gln Pro
35 40 45

55 Asp Glu Leu Leu Asp Ser Met Ser Asp Trp Cys Lys Glu His His Gly
50 55 60

Lys Ser Gly Leu Thr Lys Ala Val Lys Glu Ser Thr Ile Thr Leu Gln
65 70 75 80

60 Gln Ala Glu Tyr Glu Phe Leu Ser Phe Val Arg Gln Gln Thr Pro Pro
80 85 90

650

85 90 95
 Gly Leu Cys Pro Leu Ala Gly Asn Ser Val His Glu Asp Lys Lys Phe
 100 105 110
 5 Leu Asp Lys Tyr Met Pro Gln Phe Met Lys His Leu His Tyr Arg Ile
 115 120 125
 Ile Asp Val Ser Thr Val Lys Glu Leu Cys Arg Arg Trp Tyr Pro Glu
 130 135 140
 10 Glu Tyr Glu Phe Ala Pro Lys Lys Ala Ser His Arg Ala Leu Asp
 145 150 155 160
 15 Asp Ile Ser Glu Ser Ile Lys Glu Leu Gln Phe Tyr Arg Asn Asn Ile
 165 170 175
 Phe Lys Lys Lys Ile Asp Glu Lys Lys Arg Lys Ile Ile Glu Asn Gly
 180 185 190
 20 Glu Asn Glu Lys Thr Val Ser Xaa
 195 200

25 (2) INFORMATION FOR SEQ ID NO: 489:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 351 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 489:

Met Ala Thr Thr Ala Ala Pro Ala Gly Gly Ala Arg Asn Gly Ala Gly
 1 5 10 15
 Pro Glu Trp Gly Gly Phe Glu Glu Asn Ile Gln Gly Gly Gly Ser Ala
 20 25 30
 40 Val Ile Asp Met Glu Asn Met Asp Asp Thr Ser Gly Ser Ser Phe Glu
 35 40 45
 Asp Met Gly Glu Leu His Gln Arg Leu Arg Glu Glu Glu Val Asp Ala
 50 55 60
 45 Asp Ala Ala Asp Ala Ala Ala Glu Glu Glu Asp Gly Glu Phe Leu
 65 70 75 80
 Gly Met Lys Gly Phe Lys Gly Gln Leu Ser Arg Gln Val Ala Asp Gln
 85 90 95
 Met Trp Gln Ala Gly Lys Arg Gln Ala Ser Arg Ala Phe Ser Leu Tyr
 100 105 110
 55 Ala Asn Ile Asp Ile Leu Arg Pro Tyr Phe Asp Val Glu Pro Ala Gln
 115 120 125
 Val Arg Thr Gly Leu Leu Glu Ser Met Ile Pro Ile Lys Met Val Asn
 130 135 140
 60

651

Phe Pro Gln Lys Ile Ala Gly Glu Leu Tyr Gly Pro Leu Met Leu Val
 145 150 155 160
 Phe Thr Leu Val Ala Ile Leu Leu His Gly Met Lys Thr Ser Asp Thr
 165 170 175
 5 Ile Ile Arg Glu Gly Thr Leu Met Gly Thr Ala Ile Gly Thr Cys Phe
 180 185 190
 Gly Tyr Trp Leu Gly Val Ser Ser Phe Ile Tyr Phe Leu Ala Tyr Leu
 195 200 205
 10 Cys Asn Ala Gln Ile Thr Met Leu Gln Met Leu Ala Leu Leu Gly Tyr
 210 215 220
 Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr Tyr Asn Ile His
 225 230 235 240
 Leu His Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val Gly Gly Leu Ser
 245 250 255
 20 Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr Val Gly Pro Thr
 260 265 270
 Gln Arg Leu Leu Cys Gly Thr Leu Ala Ala Leu His Met Leu Phe
 275 280 285
 Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Val Gly Ile Leu
 290 295 300
 30 Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln Arg Val Pro Arg
 305 310 315 320
 Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro Thr Thr Val Leu
 325 330 335
 Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln Ser His Xaa
 340 345 350
 40

(2) INFORMATION FOR SEQ ID NO: 490:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 265 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 490:

Met Arg Gly Ser Arg Gly Gly Trp Ala Gly Glu Met Ala Ala Ser Gly
 1 5 10 15
 Glu Ser Gly Thr Ser Gly Gly Gly Ser Thr Glu Glu Ala Phe Met
 20 25 30
 55 Thr Phe Tyr Ser Glu Val Lys Gln Ile Glu Lys Arg Asp Ser Val Leu
 35 40 45
 Thr Ser Lys Asn Gln Ile Glu Arg Leu Thr Arg Pro Gly Ser Ser Tyr
 50 55 60

Phe Asn Leu Asn Pro Phe Glu Val Leu Glu Ile Asp Pro Glu Val Thr
 65 70 75 80
 5 Asp Glu Glu Ile Lys Lys Arg Phe Arg Glu Glu Ser Ile Leu Val His
 85 90 95
 Pro Asp Lys Asn Glu Asp Asp Ala Asp Arg Ala Glu Lys Ala Phe Glu
 100 105 110
 10 Ala Val Asp Lys Ala Tyr Lys Leu Leu Asp Glu Glu Lys Lys
 115 120 125
 Arg Ala Leu Asp Val Ile Glu Ala Gly Lys Glu Tyr Val Glu His Thr
 130 135 140
 15 Val Lys Glu Arg Lys Lys Glu Lys Lys Glu Gly Lys Pro Thr Ile
 145 150 155 160
 20 Val Glu Glu Asp Asp Pro Glu Leu Phe Lys Glu Ala Val Tyr Lys Glu
 165 170 175
 Thr Met Lys Leu Phe Ala Glu Leu Glu Ile Lys Arg Lys Glu Arg Glu
 180 185 190
 25 Ala Lys Glu Met His Glu Arg Lys Arg Glu Arg Glu Glu Ile Glu
 195 200 205
 Ala Glu Glu Lys Ala Lys Arg Glu Arg Glu Trp Glu Lys Asn Phe Glu
 210 215 220
 30 Glu Ser Arg Asp Gly Arg Val Asp Ser Trp Arg Asn Phe Glu Ala Asn
 225 230 235 240
 35 Thr Lys Gly Lys Lys Glu Lys Lys Asn Arg Thr Phe Leu Arg Pro Pro
 245 250 255
 Lys Val Lys Met Glu Glu Arg Glu Xaa
 260 265
 40
 45 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 491:
 50 Asp Ser Met Pro Thr Cys Pro Leu Xaa Ala Ser Leu Glu Cys Gly Pro
 1 5 10 15
 55 Leu Leu Pro Val Arg Leu Cys Leu
 20 25
 60 (2) INFORMATION FOR SEQ ID NO: 492:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 492:
 Phe Ser Gly Leu Ser His Leu Ser Arg Lys Phe Thr Glu Trp Cys Tyr
 65 70 75 80

(1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 159 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 493:
 5 Met Asn Glu Tyr Arg Val Pro Glu Leu Asn Val Glu Asn Gly Val Leu
 1 5 10 15
 10 Lys Ser Leu Ser Phe Leu Phe Glu Tyr Ile Gly Glu Met Gly Lys Asp
 20 25 30
 Tyr Ile Tyr Ala Val Thr Pro Leu Leu Glu Asp Ala Leu Met Asp Arg
 35 40 45
 15 Asp Leu Val His Arg Glu Thr Ala Ser Ala Val Val Glu His Met Ser
 50 55 60
 Leu Gly Val Tyr Gly Phe Gly Cys Glu Asp Ser Leu Asn His Leu Leu
 65 70 75 80
 Asn Tyr Val Trp Pro Asn Val Phe Glu Thr Ser Pro His Val Ile Glu
 85 90 95
 25 Ala Val Met Gly Ala Leu Glu Gly Leu Arg Val Ala Ile Gly Pro Cys
 100 105 110
 Arg Met Leu Glu Tyr Cys Leu Glu Gly Leu Phe His Pro Ala Arg Lys
 115 120 125
 30 Val Arg Asp Val Tyr Trp Lys Ile Tyr Asn Ser Ile Tyr Ile Gly Ser
 130 135 140
 Glu Asp Ala Leu Ile Ala His Tyr Pro Arg Ile Tyr Glu Arg Xaa
 145 150 155
 35
 40 (2) INFORMATION FOR SEQ ID NO: 493:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 279 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 493:
 45 Met Ile Ser Asp Asn Ser Ala Glu Asn Ile Ala Leu Val Thr Ser Met
 1 5 10 15
 50 Tyr Asp Gly Leu Leu Glu Ala Gly Ala Arg Leu Cys Pro Thr Val Glu
 20 25 30
 Leu Glu Asp Ile Arg Asn Leu Glu Asp Leu Thr Pro Leu Lys Leu Ala
 35 40 45
 55 Ala Lys Glu Gly Lys Ile Glu Ile Phe Arg His Ile Leu Glu Arg Glu
 50 55 60
 Phe Ser Gly Leu Ser His Leu Ser Arg Lys Phe Thr Glu Trp Cys Tyr
 65 70 75 80

Gly Pro Val Arg Val Ser Leu Tyr Asp Leu Ala Ser Val Asp Ser Cys
85 90 95

5 Glu Glu Asn Ser Val Leu Glu Ile Ile Ala Phe His Cys Lys Ser Pro
100 105 110

His Arg His Arg Met Val Val Leu Glu Pro Leu Asn Lys Leu Glu
115 120 125

10 Ala Lys Trp Asp Leu Leu Ile Pro Lys Phe Phe Leu Asn Phe Leu Cys
130 135 140

Asn Leu Ile Tyr Met Phe Ile Phe Thr Ala Val Ala Tyr His Gln Pro
145 150 155 160

Thr Leu Lys Lys Gln Ala Ala Pro His Leu Lys Ala Glu Val Gly Asn
165 170 175

20 Ser Met Leu Leu Thr Gly His Ile Leu Ile Leu Leu Gly Ile Tyr
180 185 190

Leu Leu Val Gly Gln Leu Trp Tyr Phe Trp Arg Arg His Val Phe Ile
195 200 205

25 Trp Ile Ser Phe Ile Asp Ser Tyr Phe Glu Ile Leu Phe Leu Gln
210 215 220

Ala Leu Leu Thr Val Val Ser Gln Val Leu Cys Phe Leu Xaa Ile Glu
225 230 235 240

Trp Tyr Leu Pro Leu Leu Val Ser Ala Leu Val Leu Gly Trp Leu Asn
245 250 255

35 Leu Leu Tyr Tyr Thr Arg Gly Phe Gln His Thr Gly Ile Tyr Ser Val
260 265 270

Met Ile Gln Lys Pro Trp Xaa
275

40

(2) INFORMATION FOR SEQ ID NO: 494:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 193 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:

Met Ile Arg Cys Gly Leu Ala Cys Glu Arg Cys Arg Trp Ile Leu Pro
1 5 10 15

Leu Leu Leu Leu Ser Ala Ile Ala Phe Asp Ile Ile Ala Leu Ala Gly
20 25 30

Arg Gly Trp Leu Gln Ser Ser Asp His Gly Gln Thr Ser Ser Leu Trp
35 40 45

60 Trp Lys Cys Ser Gln Glu Gly Gly Ser Gly Ser Tyr Glu Glu Gly

50 55 60

Cys Gln Ser Leu Met Glu Tyr Ala Trp Gly Arg Ala Ala Ala Met
65 70 75 80

Leu Phe Cys Gly Phe Ile Ile Leu Val Ile Cys Phe Ile Leu Ser Phe
85 90 95

10 Phe Ala Leu Cys Gly Pro Gln Met Leu Val Phe Leu Arg Val Ile Gly
100 105 110

Gly Leu Leu Ala Leu Ala Val Phe Gln Ile Ile Ser Leu Val Ile
115 120 125

15 Tyr Pro Val Lys Tyr Thr Gln Thr Phe Thr Leu His Ala Asn Xaa Ala
130 135 140

Val Thr Tyr Ile Tyr Asn Trp Ala Tyr Gly Phe Gly Trp Ala Thr
145 150 155 160

20 Ile Ile Leu Ile Gly Cys Ala Phe Phe Cys Cys Leu Pro Asn Tyr
165 170 175

Glu Asp Asp Leu Leu Gly Asn Ala Lys Pro Arg Tyr Phe Tyr Thr Ser
180 185 190

Ala

25

30

(2) INFORMATION FOR SEQ ID NO: 495:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 205 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 495:

Met Ala Ala Gly Asp Gln Val Phe Ser Gly Ala Gly His Val Xaa Glu
1 5 10 15

His Val Ala Gly Arg His Ala Trp Leu Leu Thr Trp Gln Ser Ala
20 25 30

Cys Pro Ala Asn Arg Leu Ser Leu Val Pro Leu Val Pro Ser Ala Ser
35 40 45

Met Thr Arg Leu Met Arg Xaa Arg Thr Ala Ser Gly Ser Ser Val Ile
50 55 60

Leu Trp Met Ala Pro Ala Ala Pro Thr Pro Ala Arg Ala Pro Glu
65 70 75 80

55 Ala Ala Pro Thr Pro Ala Arg Ala Pro Ala Ala Arg Thr Pro Ala
85 90 95

Arg Gly Pro Thr Trp Thr Ser Pro Thr Arg Val Leu Leu Gly Thr
100 105 110

Xaa Pro Gly Pro Ser Pro Trp Arg Ser Pro Ala Arg Arg Pro Ala Gln
115 120 125

5 Leu Pro Pro Asp Ser Asp Leu Cys Ser Gly Pro Leu Leu Pro Gly
130 135 140

Pro Phe Ser Pro Pro Ala Cys His Thr Ala Pro Ann Ser Val Leu Ile
145 150 155 160

10 Gln Ser Leu Phe Cys Lys Ser Glu Leu Trp Trp Arg Gln Met Arg Ser
165 170 175

Ile Thr Trp Val Pro Ser Pro Lys Ala Gly Trp Arg Trp Thr Lys Gly
180 185 190

15 Arg Lys Gln Ala Ser Pro His Arg Ile Leu Phe His Xaa
195 200 205

20 (2) INFORMATION FOR SEQ ID NO: 496:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 147 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 496:

25 Met Ala Leu Thr Leu Leu Pro Ser Val Ser Arg Leu Pro Gly Glu Arg
1 5 10 15

30 Met Ala Ala Ser Gly Leu Pro Tyr Val Leu His His Lys Ser Ser Leu
20 25 30

35 Met Lys Val Ile Phe Phe Pro Tyr Pro Val Leu Pro Leu Pro Ala Pro
35 40 45

Ann Gly Thr Trp Val Pro Arg Leu Val Leu Gly Leu Gly Ser Gly Asp
50 55 60

40 Gln Val His Tyr Leu Pro Ile Ser Ser Ile Val Asn Tyr Gly Thr
65 70 75 80

Ser Val Ser Gly Lys Ser Trp Val Phe Leu Val Tyr Pro Leu His Pro
85 90 95

Thr Pro Thr Trp Ser Thr Arg Cys Phe Gln Val Trp Asp Leu Leu Ser
100 105 110

50 Val Glu Leu Pro Asp Lys Gly Glu Gly Asn Thr Arg Arg Ala Ser Gly
115 120 125

Val Pro Gly Leu Ser Gln Leu Pro Thr Ser His Lys Pro Ile Lys Gln
130 135 140

Glu Tyr Xaa

145

60

(2) INFORMATION FOR SEQ ID NO: 497:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 497:

5 Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val
1 5 10 15

10 Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys Glu Val Gly
20 25 30

15 Ala Ala Leu Pro Pro Arg Gly Pro Ser Leu Ser Asp Cys Leu Gly Leu
35 40 45

Pro Pro Trp Thr Pro Trp Gly Pro Ala Trp Thr Leu Ala Gln Ser Xaa
50 55 60

25 (2) INFORMATION FOR SEQ ID NO: 498:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 94 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 498:

35 Met Ser Thr Gly Ala Leu Ann Thr Ser Pro Pro Ala Ser Ann Arg Leu
1 5 10 15

Glu Ser Thr Leu Ann Glu Tyr Leu Ile Gln Pro Gln Leu His Cys Ser
20 25 30

40 Ser Val Gln Arg Leu Thr Leu Lys Trp Gly Cys Ser Ser Leu Gln Arg
35 40 45

Asp Gly Gln Ala Val Pro Trp Gly Leu Trp Gln Arg Ala Tyr Pro Ser
50 55 60

Leu Leu Pro Thr Leu Pro Ser Asp Leu Leu Arg Pro His Ala Val Thr
65 70 75 80

Pro Ser Val Ser Val Ser Val His Thr Cys Glu Ser Ser Xaa
85 90

(2) INFORMATION FOR SEQ ID NO: 499:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 499:

Glu Tyr Xaa

145

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Xaa Leu Phe Ser Ser Ser
1 5 10 15

5 Leu Pro Phe Leu Trp Leu
20

10 (2) INFORMATION FOR SEQ ID NO: 500:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 500:

Arg Gly Gly Leu Cys Pro Leu Leu Val Pro Gly Pro Leu Ala Arg Gln
1 5 10 15

20 Glu Pro Ser Pro Ser Leu Gln Gly Cys Ser Glu Ser Pro Val Gly Met
25

25 Asp

30 (2) INFORMATION FOR SEQ ID NO: 501:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 501:

Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu
1 5 10 15

40 Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser
20 25

45 (2) INFORMATION FOR SEQ ID NO: 502:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 502:

Pro Gly Lys Pro Gln Ala Cys Pro Glu Leu Thr Ser Val Leu Pro
1 5 10 15

55 (2) INFORMATION FOR SEQ ID NO: 503:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

60

(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 503:

5 Asn Lys Ser Leu Xaa Ser Cys Leu Phe Val Leu His Phe Val Leu His
1 5 10 15

10 Cys Xaa Phe

15 (2) INFORMATION FOR SEQ ID NO: 504:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 504:

20 Met Glu Lys Thr His Arg Leu Arg Ile Arg Asn Pro Cys Leu Gln Phe
1 5 10 15

25 Ser Ile Leu Asn Leu Phe Leu Leu Lys Met Ile Val Ser
20 25

30 (2) INFORMATION FOR SEQ ID NO: 505:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 75 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 505:

35 Met Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln
1 5 10 15

40 Asn Leu Ser Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu
20 25 30

45 Ala Gly Lys Leu Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly
35 40 45

Pro Ser Ser Phe Gln Asn Pro Ala Ser Ser Pro Ser Ser Trp Thr His
50 55 60

50 Glu Glu Glu Pro Gly Tyr Phe Pro Gln Tyr Xaa
65 70 75

55 (2) INFORMATION FOR SEQ ID NO: 506:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 506:

Leu Pro Leu Ala Glu Leu Lys Asn Trp Val
1 5 10

5

(2) INFORMATION FOR SEQ ID NO: 507:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 507:

Met Leu Trp Phe Gly Gly Cys Ser Ala Val Asn Ala Thr Gly His Leu
1 5 10 15

Ser Asp Thr Leu Trp Leu Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr
20 25 30

Gly Asp Val Val Pro Gly Thr Met Trp Gly Lys Ile Val Cys Leu Cys
35 40 45

Thr Gly Val Met Gly Val Cys Cys Thr Ala Leu Leu Val Ala Val
50 55 60

Ala Arg Lys Leu Glu Phe Asn Lys Ala Glu Lys His Val His Asn Phe
65 70 75 80

Met Met Asp Ile Gln Tyr Thr Lys Glu Met Lys Glu Ser Ala Ala Arg
85 90 95

Val Leu Gln Glu Ala Trp Met Phe Tyr Lys His Thr Arg Arg Lys Glu
100 105 110

Ser His Ala Ala Arg Arg His Gln Arg Xaa Leu Leu Ala Ala Ile Asn
115 120 125

Ala Phe Arg Gln Val Arg Leu Lys His Arg Lys Leu Arg Glu Gln Val
130 135 140

Asn Ser Met Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu
145 150 155 160

Gln Gln Asn Leu Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp
165 170 175

Thr Leu Ala Gly Lys Leu Asp Ala Leu Thr Gly Leu Leu Ser Thr Ala
180 185 190

Leu Gly Pro Arg Gln Leu Pro Glu Pro Ser Gln Gln Ser Lys Xaa
195 200 205

55

(2) INFORMATION FOR SEQ ID NO: 508:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids

60

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 508:

Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val
1 5 10 15

Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala
20 25 30

Val Xaa Lys Lys
35

15

(2) INFORMATION FOR SEQ ID NO: 509:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 509:

Met Ala Leu Val Ala Leu Phe Thr Gln Leu Met Arg Xaa Leu Gly Arg
1 5 10 15

Cys Pro Gln

30

(2) INFORMATION FOR SEQ ID NO: 510:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 510:

Met Thr Phe Pro Phe Glu Lys Glu Asn Ser Cys Phe Gln Cys Leu Leu
1 5 10 15

Phe Asp Ser Trp Arg Glu Gln Thr Arg Thr Asn Ile Gln Pro Gln Arg
20 25 30

45

50

(2) INFORMATION FOR SEQ ID NO: 511:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 511:

Met His Leu Leu Asp Phe Phe Arg Asp Leu Val Leu Leu Val Leu Leu
1 5 10 15

662

663

Ala Leu Leu Asp Ser Phe Trp Leu Glu Val Gln Lys
20 25

5

(2) INFORMATION FOR SEQ ID NO: 512:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 512:

10

Met Cys Leu Ile His Phe Ile Lys Ile Ile Leu Val Phe Ile Leu Lys
1 5 10 15

Leu Trp Leu Tyr Ser Gln Lys Cys Pro Lys
20 25

20

(2) INFORMATION FOR SEQ ID NO: 513:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 513:

25

Met Ile His Val His Glu Trp Asn Asp Gln Met Leu Met Val Tyr Ile
1 5 10 15

Phe Leu Tyr Pro Val Ser Ile Thr Phe Leu Asn Leu Cys Ser Leu Thr
20 25 30

Cys

40

(2) INFORMATION FOR SEQ ID NO: 514:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

45

Leu Asn Glu Ser Tyr Val Ser Arg Ala Gly Gly Trp Phe Ser Met Phe
1 5 10 15

Xaa Leu Ile Phe Phe Leu Leu Ala Leu Gly Ser Xaa Leu Cys Leu Leu
20 25 30

55

Leu Cys Leu Pro Ser Phe Asn Lys Thr Arg Lys Gln Lys Pro
35 40 45

60

(2) INFORMATION FOR SEQ ID NO: 515:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 515:

5

Ser Ser Lys Thr Pro Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly Ser
1 5 10 15

Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu Gly
20 25 30

Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp
35 40

(2) INFORMATION FOR SEQ ID NO: 516:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 516:

25

Leu Asn Trp
1

30

(2) INFORMATION FOR SEQ ID NO: 517:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 517:

35

Phe Ala Phe Cys Ala Glu Leu Met Ile Gln Asn Trp Thr Leu Gly Ala
1 5 10 15

Val Asp Ser Gln Met Asp Asp Met Asp Met Asp Leu Asp Lys Glu Phe
20 25 30

Leu Gln Asp Leu Lys Glu Leu Lys Val Leu Val Ala Asp Lys Asp Leu
35 40 45

Leu Asp Leu His Lys Ser Leu Val Cys Thr Ala Leu Arg Gly Lys Leu
50 55 60

Gly Val Phe Ser Glu Met Glu Ala Asn Phe Lys Asn Leu Ser Arg Gly
65 70 75 80

Leu Val Asn Val Ala Ala Lys Leu Thr His Asn Lys Asp Val Arg Asp
85 90 95

Leu Phe Val Asp Leu Val Glu Lys Phe Val Glu Pro Cys Arg Ser Asp
100 105 110

His Trp Pro Leu Ser Asp Val Arg Phe Leu Asn Gln Tyr Ser Ala
115 120 125

5 Ser Val His Ser Leu Asp Gly Phe Arg His Gln Ala Ser Gly Thr Ala
130 135 140

Thr Trp Ala Pro Ser Ala Ala Ser Cys Ala Cys Ile Met Thr Glu
145 150 155

10 Val Pro Pro Asn Ala Pro Thr Leu Thr Ile Lys Leu Leu
165 170

(2) INFORMATION FOR SEQ ID NO: 518:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 518:

Met Trp Lys Asn Leu Leu Gly Ser Gly Ser Val Phe Val Thr Trp Phe Ser
1 5 10 15

Leu Val Met Ile Leu Ser Gly Ile Gly Pro Leu Gly Asp Ala Glu Asp
20 25 30

30 Ser Ile Ser Asp Val Ser His Arg Leu Arg Pro
35 40

35 (2) INFORMATION FOR SEQ ID NO: 519:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 519:

Phe Gln Phe Pro Leu Leu Thr Ile Ala Leu Gln Phe Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO: 520:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 520:

Met His Tyr Val Ile Val Leu Ser Leu Phe Val Val Leu Glu Lys Lys
1 5 10 15

Asn Lys Met Gly Ser Asp Gly Cys Leu Arg Lys Asn Gly Ser
20 25 30

(2) INFORMATION FOR SEQ ID NO: 521:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 521:

Met Ser Arg Ser Ile Val Leu Arg Gly Ser Leu Phe Leu Phe Phe Ser
1 5 10 15

15 His Tyr Thr Leu Lys Leu Ser Val Ile Lys Gln Thr Asn Arg Lys
20 25 30

Ile Val Trp Glu Lys Pro Cys Ile Arg Leu Phe Tyr Xaa Val Leu
35 40 45

(2) INFORMATION FOR SEQ ID NO: 522:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:

Met Pro Leu Pro Val Leu Leu Cys Leu Thr Leu Pro Met Pro Leu Pro
1 5 10 15

Ser Ala Thr Ala Arg Gly Asn Arg Thr
20 25

(2) INFORMATION FOR SEQ ID NO: 523:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 523:

Ser Ser Ile Pro Val Ser Ile Leu Ile Gly Met Lys Leu Ile Leu Tyr
1 5 10 15

Leu Leu Ile Thr Glu Ser Gly Ser His Glu Lys Lys Ser Phe Tyr Pro
20 25 30

Ser Phe Lys Tyr Met Phe Lys Ile Ile Ile Tyr Val Ser Ala Tyr Cys
35 40 45

Arg Thr Ala Leu Arg Ala Thr Val Ser His
50 55

60

(2) INFORMATION FOR SEQ ID NO: 524:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 524:

5 Asn Arg Thr Leu Leu Phe Ile Leu Phe Val Leu Phe Gly Leu Gly 15
1 5 10

Tyr Gly Phe

15

(2) INFORMATION FOR SEQ ID NO: 525:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 525:

25 Met Phe Leu Leu Val Leu Ser Val Phe Cys Asp Phe Met Cys Ser Ile 15
1 5 10

Ala Pro Arg Cys His Ala Leu Ser Leu Val Ser Leu Arg Ala Gln His 30
20 25

30 Leu Ser Leu Phe Ile Thr Cys His 40
35

35

(2) INFORMATION FOR SEQ ID NO: 526:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 526:

45 Met Leu Leu Phe Ile Leu Leu Thr Leu Ser Ser Gly Cys Arg Leu Leu 15
1 5 10

Val Ser Ser Trp Lys Thr Phe Leu Pro His Ser Leu Pro Gly Pro 30
20 25

50 Arg Glu His Pro Glu Gly Ser Arg Thr Trp Phe Arg Tyr Trp Glu 45
35 40

Pro Gly Ala His Cys Leu His Cys Ala 55
50 55

55

(2) INFORMATION FOR SEQ ID NO: 527:

- (1) SEQUENCE CHARACTERISTICS:

60 Ser Pro Gly Leu Lys Asn Gly Ile Phe Leu Phe Leu Arg Ala Met 110
100 105

- (A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 527:

5 Ala Arg Leu Leu Leu Phe Leu Ser Ser Val His Pro Ser Ile Met Pro 15
1 5 10

10 Ser Cys Asn Gln Leu 20

(2) INFORMATION FOR SEQ ID NO: 528:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 528:

20 Met Ser Leu Thr Ser Ser Leu Thr Phe Leu Ser His Ile Leu Leu Leu 15
1 5 10

25 Pro Gln Lys Leu Gln Phe Leu Ser Trp Met Glu Arg Gln Gln Arg Cys 30
20 25

30 Thr Gly Val Ala Lys Tyr Ala 35

(2) INFORMATION FOR SEQ ID NO: 529:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 128 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 529:

40 Met Val Leu Arg Leu Ile Gln Leu Ile Phe Leu Ile Phe Ile His 15
1 5 10

45 Ile Ile Ile Leu Leu Ile Pro Gly Ser Arg Pro Cys Gly Ser Trp Val 30
20 25

Asn Asp Arg Xaa Leu Gly Leu Arg Asp Val Thr His Leu Ile Tyr Leu 45
35 40

50 His Trp Val His Gly His Leu Pro Trp Cys His Pro Tyr Ile Gln Val 60
50 55

Glu Phe Ser Ala Leu Ile Glu Ser Thr Ala Gln Leu Gly Leu Pro Phe 80
65 70 75

55 Ser Trp Val Arg Val Ile His Pro Phe Leu Val Leu Pro Cys Leu Tyr 95
85 90

60 Ser Pro Gly Leu Lys Asn Gly Ile Phe Leu Phe Leu Arg Ala Met 110
100 105

Pro Gly Gly Met Phe Pro Gly Asn Leu Glu Ala Phe Arg Val Pro Val
115 120 125

5

10 (2) INFORMATION FOR SEQ ID NO: 530:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 82 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 530:

Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser
1 5 10 15

20 Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe
20 25 30

25 Trp Gly Phe Gly Xaa Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr
35 40 45

Pro Pro Asp Gln Glu Pro Gln Lys Thr Glu Thr Pro Ala Thr Ser Ser
50 55 60

30 His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln
65 70 75 80

Pro Asn

35

(2) INFORMATION FOR SEQ ID NO: 531:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 531:

Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala
1 5 10 15

Tyr Trp Thr Met
20

50

(2) INFORMATION FOR SEQ ID NO: 532:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 75 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 532:

55

60 (2) INFORMATION FOR SEQ ID NO: 533:

Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Gln Leu Lys Ile
1 5 10 15

5 Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu Lys Ile Ile
20 25 30

10 Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Ser Trp Ala Ile
35 40 45

Lys Ala Gln Leu Lys Ile Glu Asn Lys Asp Leu Asn Lys Thr Ala
50 55 60

15 Lys Gly Gly Gly Gln Glu Ala Lys Thr Cys Thr
65 70 75

(2) INFORMATION FOR SEQ ID NO: 533:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 60 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 533:

Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser
1 5 10 15

30 Phe Ile Val Thr Pro Thr Ser Thr Ser Asn Thr Xaa Ser Tyr Leu Trp
20 25 30

Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Xaa Trp Ala Cys
35 40 45

35 Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu
50 55 60

(2) INFORMATION FOR SEQ ID NO: 534:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 534:

Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe
1 5 10 15

Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg
20 25 30

55 Ile Leu Phe Phe Ile Val Phe
35

670

671

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 535:

5

Met Leu

1

10

- (2) INFORMATION FOR SEQ ID NO: 536:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 536:

15

Met Asp Gln Phe Lys Ile Phe Tyr Phe Leu Lys Ala Phe Phe Ala Cys

1

20

Cys Asn Val Gln Asp Pro Ser Pro Phe Met Gly Glu Thr Gly Ser Tyr

20

25

Leu Asn Ile Gly

35

30

- (2) INFORMATION FOR SEQ ID NO: 537:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 537:

35

Met Phe Asp Phe Leu Ser Tyr Phe Lys Asp Leu Leu Ser Cys

1

40

- (2) INFORMATION FOR SEQ ID NO: 538:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 538:

45

Met Gly Phe Gly Phe Val Leu Asn Ile Phe Ser Phe Phe Leu Xaa Pro

1

55

Pro Leu

60

- (2) INFORMATION FOR SEQ ID NO: 539:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 539:

5

Leu Leu Leu Trp Thr Leu Leu Ala Xaa Tyr Xaa

1

10

- (2) INFORMATION FOR SEQ ID NO: 540:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 108 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 540:

15

Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu

1

20

Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg

25

25

Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe

35

30

Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Leu Cys

50

35

Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe

65

35

Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu

85

40

Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Xaa

100

45

- (2) INFORMATION FOR SEQ ID NO: 541:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 106 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 541:

45

Phe Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met

1

55

Leu Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Tyr Ile Leu Tyr

20

60

Leu Arg Thr Leu Glu Ser Lys Leu Val Leu Phe Gly Arg Glu Val Ser

35

60

Pro Ala His Gln Tyr Ala Leu Ala Gly Ile Ser Phe Pro Phe Phe
50 55 60

5 Trp Leu Ala Gly Ala Gly Ser Ala Val Phe Trp Val Leu Gly Ala Thr
65 70 75 80

Leu Val Val Ile Gly Ser His Ala Ala Phe His Gln Ile Glu Ala Val
85 90 95

10 Asp Gly Glu Glu Leu Gln Met Glu Pro Val
100 105

15 (2) INFORMATION FOR SEQ ID NO: 542:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 116 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 542:

Met Asp Arg Phe Thr Val Ala Gly Val Leu Pro Asp Ile Glu Gln Phe
1 5 10 15

Phe Asn Ile Gly Asp Ser Ser Gly Leu Ile Gln Thr Val Phe Ile
20 25 30

30 Ser Ser Tyr Met Val Leu Ala Pro Val Phe Gly Tyr Leu Gly Asp Arg
35 40 45

Tyr Asn Arg Lys Tyr Leu Met Cys Gly Gly Ile Ala Phe Trp Ser Leu
50 55 60

Val Thr Leu Gly Ser Ser Phe Ile Pro Gly Glu His Phe Trp Leu Leu
65 70 75 80

Leu Leu Thr Arg Gly Leu Val Gly Val Gly Glu Ala Ser Tyr Ser Thr
85 90 95

Ile Ala Pro Thr Leu Ile Ala Asp Leu Phe Val Ala Asp Gln Arg Thr
100 105 110

45 Gly Cys Ser Ala Ser Thr Leu Pro Phe Arg Trp Ala Val Val Trp
115 120 125

Ala Thr Leu Gln Ala Pro Lys Xaa
130 135

50 (2) INFORMATION FOR SEQ ID NO: 543:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 424 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 543:

Met Ala Gly Asp Trp His Trp Ala Leu Arg Val Thr Pro Gly Leu Gly
1 5 10 15

5 Val Val Ala Val Leu Leu Phe Leu Val Val Arg Glu Pro Pro Arg
20 25 30

Gly Ala Val Glu Arg His Ser Asp Leu Pro Leu Asn Pro Thr Ser
35 40 45

10 Trp Trp Ala Asp Leu Arg Ala Leu Ala Arg Asn Pro Ser Phe Val Leu
50 55 60

Ser Ser Leu Gly Phe Thr Ala Val Ala Phe Val Thr Gly Ser Leu Ala
65 70 75 80

15 Leu Trp Ala Pro Ala Phe Leu Leu Arg Ser Arg Val Val Leu Gly Glu
85 90 95

20 Thr Pro Pro Cys Leu Pro Gly Asp Ser Cys Ser Ser Ser Asp Ser Leu
100 105 110

Ile Phe Gly Leu Ile Thr Cys Leu Thr Gly Val Leu Gly Val Gly Leu
115 120 125

25 Gly Val Glu Ile Ser Arg Arg Xaa Arg His Ser Asn Pro Arg Ala Asp
130 135 140

30 Pro Leu Val Cys Ala Thr Gly Leu Leu Gly Ser Ala Pro Phe Leu Phe
145 150 155 160

Leu Ser Leu Ala Cys Ala Arg Gly Ser Ile Val Ala Thr Tyr Ile Phe
165 170 175

35 Ile Phe Ile Gly Glu Thr Leu Leu Ser Met Asn Trp Ala Ile Val Ala
180 185 190

Asp Ile Leu Leu Tyr Val Val Ile Pro Thr Arg Arg Ser Thr Ala Glu
195 200 205

40 Ala Phe Gln Ile Val Leu Ser His Leu Leu Gly Asp Ala Gly Ser Pro
210 215 220

Tyr Leu Ile Gly Leu Ile Ser Asp Arg Leu Arg Asn Trp Pro Pro
225 230 235 240

Ser Phe Leu Ser Glu Phe Arg Ala Leu Gln Phe Ser Leu Met Leu Cys
245 250 255

50 Ala Phe Val Gly Ala Leu Gly Gly Ala Leu Ser Trp Ala Pro Xaa Ser
260 265 270

Ser Leu Arg Pro Thr Ala Gly Gly His Ser Cys Thr Cys Arg Ala Cys
275 280 285

55 Cys Thr Lys Gln Gly Pro Gln Thr Thr Gly Leu Trp Cys Pro Ser Gly
290 295 300

Ala Ala Pro Ala Cys Pro Trp Pro Val Cys Ser Ser Glu Arg Leu
305 310 315 320

60

674

Pro Leu Thr Tyr Leu His Ile Cys His Ser Xaa Pro Trp Ala His Pro
325 330 335

Thr Lys Gly Leu Gly Leu Thr Pro Trp Pro Gly Pro Ala Ser Arg Gly
340 345 350

Thr Leu Gly Arg Val Pro Ala Pro Arg His Tyr Xaa Gly Ser Ser Gly
355 360 365

Glu Glu Val Gly Val Gln Glu Gly Asp Pro Ser Pro Gln Gly Xaa Pro
370 375 380

Gln Gly Leu Gly Ala Ile Cys Asn Gly Ile Lys Phe Val Ala Arg Pro
385 390 395 400

Gln Val Pro Ala Leu Val Phe Leu Trp Val Ala Ser Asp Leu Ala Pro
405 410 415

Arg Leu His Pro Arg Ala Pro Glu
420

(2) INFORMATION FOR SEQ ID NO: 544:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 544:

Met Phe Arg Phe Val Ile Cys Leu Phe Leu Trp Leu Val Leu Cys Arg
1 5 10 15

Asp Ser Thr Ser Ala Ser Arg Ile Ala Leu Tyr Tyr Arg Ile Val Phe
20 25 30

Leu Ile His Gln Cys Ser Ser
35

(2) INFORMATION FOR SEQ ID NO: 545:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 545:

Met Leu Pro Trp Xaa Ala Gln Leu Leu Asp Arg Thr Ile Gly Pro Leu
1 5 10 15

Tyr Leu Leu Phe Val Gln Phe Ser Pro Ala Phe Ser Arg Thr Ser Pro
20 25 30

Trp Arg Ser Pro Lys Asn Phe Arg Arg Leu Tyr Pro Pro Cys Thr Thr
35 40 45

Ser Gly Cys Ala Ala Arg Trp Leu Phe Ser

675

WO 98/39448

50 55

5 (2) INFORMATION FOR SEQ ID NO: 546:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 546:

Met Gly Leu Ser Val Leu Leu Pro Leu Cys Leu Leu Gly Pro Gly Arg
1 5 10 15

Phe Thr Ser Gly Gln Lys Pro Leu Asp Thr Pro Gly Leu Gly Val Pro
20 25 30

Phe

25 (2) INFORMATION FOR SEQ ID NO: 547:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 367 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 547:

Met Ala Lys Pro Gln Val Val Ala Pro Val Leu Met Ser Lys Leu
1 5 10 15

Ser Val Asn Ala Pro Glu Phe Tyr Pro Ser Gly Tyr Ser Ser Ser Tyr
20 25 30

Thr Glu Ser Tyr Glu Asp Gly Cys Glu Asp Tyr Pro Thr Leu Ser Glu
35 40 45

Tyr Val Gln Asp Phe Leu Asn His Leu Thr Glu Gln Pro Gly Ser Phe
50 55 60

Glu Thr Glu Ile Glu Gln Phe Ala Glu Thr Leu Asn Gly Cys Val Thr
65 70 75 80

Thr Asp Asp Ala Leu Gln Glu Leu Val Glu Leu Ile Tyr Gln Gln Ala
85 90 95

Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr
100 105 110

Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln
115 120 125

Leu Leu Leu Gln Arg Cys Arg Thr Glu Tyr Glu Val Lys Asp Gln Ala
130 135 140

Ala Lys Gly Asp Glu Val Thr Arg Lys Arg Phe His Ala Phe Val Leu
145 150 155 160

Phe Leu Gly Glu Leu Tyr Leu Asn Leu Glu Ile Lys Gly Thr Asn Gly
 165 170 175
 5 Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Glu Leu Leu
 180 185 190
 Asn Ala Leu Phe Ser Asn Pro Met Asp Asp Asn Leu Ile Cys Ala Val
 195 200 205
 10 Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Glu Asp Ala Trp Lys Glu
 210 215 220
 Lys Gly Lys Met Asp Met Glu Glu Ile Ile Gln Arg Ile Glu Asn Val
 225 230 235 240
 Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Lys
 245 250 255
 15 Leu Val Glu Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser
 260 265 270
 Thr Tyr Arg Glu Ala Thr Pro Glu Asn Asp Pro Asn Tyr Phe Met Asn
 275 280 285
 20 Glu Pro Thr Phe Thr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp
 290 295 300
 Pro Asp Tyr Gln Glu Lys Tyr Gln Glu Leu Leu Glu Arg Glu Asp Phe
 305 310 315 320
 Phe Pro Asp Tyr Glu Glu Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp
 325 330 335
 35 Pro Tyr Leu Asp Asp Ile Asp Asp Glu Met Asp Pro Glu Ile Glu Glu
 340 345 350
 Ala Tyr Glu Lys Phe Cys Leu Glu Ser Glu Arg Lys Arg Lys Gln
 355 360 365
 40
 (2) INFORMATION FOR SEQ ID NO: 548:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 77 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 548:
 Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met
 1 5 10 15
 Leu Ile Val Ser Val Leu Ala Leu Ile Pro Glu Thr Thr Thr Thr
 20 25 30
 Val Gly Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Leu Ala
 35 40 45
 60 Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro

Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu
 65 70 75
 5
 (2) INFORMATION FOR SEQ ID NO: 549:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 47 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 549:
 Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser
 1 5 10 15
 His Cys Trp Gly Leu Pro Leu Ala Cys Gly Thr Phe Val Gln Gly His
 20 25 30
 Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala
 35 40 45
 25
 (2) INFORMATION FOR SEQ ID NO: 550:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 168 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 550:
 Met Leu Leu Ser Leu Ala Ala Phe Ser Val Ile Ser Val Val Ser Tyr
 1 5 10 15
 Leu Ile Leu Ala Leu Ser Val Thr Ile Ser Phe Arg Ile Tyr Lys
 20 25 30
 Ser Val Ile Gln Ala Val Gln Lys Ser Glu Glu Gly His Pro Phe Lys
 35 40 45
 Ala Tyr Leu Asp Val Asp Ile Thr Leu Ser Ser Glu Ala Phe His Asn
 50 55 60
 Tyr Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile
 65 70 75 80
 Ile Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala
 85 90 95
 Val Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile
 100 105 110
 Thr Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val
 115 120 125
 Tyr Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg
 130 135 140

Asp Gln Thr Lys Ser Ile Val Glu Lys Ile Gln Ala Lys Leu Pro Gly
145 150 155 160

5 Ile Ala Lys Lys Ala Glu Xaa
165

10 (2) INFORMATION FOR SEQ ID NO: 551:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 124 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 551:

Ser Val Pro Phe His Leu Leu Val Val Leu Arg Ser Arg Ala Val Arg
1 5 10 15

20 Ala Arg Arg Arg Arg Glu Pro Arg Ser Leu Pro Arg Pro Gly Asp Glu
20 25 30

25 Glu Leu Gln Leu Leu Cys Gly Ala Arg Ser Asp Phe Leu Glu Arg
35 40 45

Cys Glu Glu Asp Trp Val Cys Leu Trp His His Ala Asp His Ala Ala
50 55 60

30 Phe Pro Gly Ser Phe Gln Cys His Gln Cys Gly Phe Leu Pro His Pro
65 70 75 80

35 Gly Ser Ser Leu Cys His His Gln Leu Gln Asp Leu Gln Val Arg His
85 90 95

Pro Ser Cys Thr Glu Val Arg Arg Arg Pro Ser Ile Gln Ser Leu Pro
100 105 110

40 Gly Arg Arg His Tyr Ser Val Leu Arg Ser Phe Pro
115 120

(2) INFORMATION FOR SEQ ID NO: 552:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 552:

Met Val His Leu Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp
1 5 10 15

55 Val Thr Phe Val Ser Gly Phe Leu Phe Arg Ser Leu Pro Arg His
20 25 30

Thr Phe Gly Leu Val Gln Ser Lys Leu Phe Pro Phe Tyr Phe His Ile
35 40 45

Ser Met Gly Cys Ala Phe Ile Asn Leu Cys Ile Leu Ala Ser Gln His
50 55 60

5 Ala Trp Ala Gln Leu Thr Phe Trp Glu Ala Ser Gln Leu Tyr Leu Leu
65 70 75 80

Phe Leu Ser Leu Thr Leu Ala Thr Val Asn Ala Arg Trp Leu Glu Pro
85 90 95

10 Arg Thr Thr Ala Ala Met Trp Ala Leu Gln Thr Val Glu Lys Glu Arg
100 105 110

Gly Leu Gly Gly Glu Val Pro Gly Ser His Gln Gly Pro Asp Pro Tyr
115 120 125

15 Arg Gln Leu Arg Glu Lys Asp Pro Lys Tyr Ser Ala Leu Arg Gln Asn
130 135 140

20 Phe Phe Arg Tyr His Gly Leu Ser Ser Leu Cys Asn Leu Gly Cys Val
145 150 155 160

Leu Ser Asn Gly Leu Cys Leu Ala Gly Leu Ala Leu Glu Ile Arg Ser
165 170 175

25 Leu

30 (2) INFORMATION FOR SEQ ID NO: 553:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 553:

Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu
1 5 10 15

40 Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His
20 25 30

45 Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys
35 40 45

Thr Glu Gly Ile Pro Cys Ala Lys Ile Pro Glu Trp Val Thr His Leu
50 55 60

50 Thr Trp Gln Thr Leu Lys Asn Ser
65 70

55 (2) INFORMATION FOR SEQ ID NO: 554:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

680

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 554:

5 Val Leu Arg Ile Ile Cys Leu Trp Pro Cys Gly Thr Thr Leu Pro Leu 15
1 5 10
Val Glu Lys Ala His Asp Ser His Ser Ala Asp Pro Val Cys Pro Gly 30
20 25
Leu Thr Ala His Leu Pro Val Leu Leu Tyr Val Gln Leu 45
35 40

(2) INFORMATION FOR SEQ ID NO: 555:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 251 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 555:

20 Met Lys His Ala Asp Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser 15
1 5 10

25 Pro Leu Leu Met Thr Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu 30
20 25

Ser Leu Gly Pro Arg Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg 45
35 40

Gly Phe Met Ile Val Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr 60
50 55

35 Ile Val Tyr Glu Phe Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp 75
65 70 75 80

Arg Cys Asp Pro Gln Asp Cys Thr Leu Gly Gln Cys Pro Ser Val Pro 95
85 90

40 Ser Pro Xaa Thr Pro Val Thr Lys Ala Tyr Val Val Arg Thr Glu Gln 110
100 105 110

Gly Thr Gly Pro Pro Leu Pro Thr Ala Ala Leu Gln Gly Pro Arg Leu 125
115 120 125

45 Trp Phe Leu Thr His Phe Pro Arg Ala Ala Pro Gly Met Trp Pro His 140
130 135 140

50 Cys Cys Leu Pro Leu Gln Ser Trp Gly Leu Lys Gly Leu Tyr Ser Tyr 155
145 150 155 160

Phe Pro Leu Pro Ala Leu Lys Leu Gly Arg Gly Ala Leu Arg Ala Gly 175
165 170 175

55 Pro Thr Lys Gly Leu Val Ala Phe Phe Leu Thr Gln Lys Arg Ser Ala 185
180 185 190

Ile Met Ser Leu Trp Thr Gln Ser His Ser Thr Thr Pro His Thr Glu 205
195 200 205

60

681

Ala Val Ala Ser Gly Pro Lys Val Arg Val Gly Gly Leu Gly Ile 220
210 215 220

5 Gln Pro Val Glu Ala Ala Tyr Ser Thr Cys Val Leu Ile Lys Ser Asp 240
225 230 235 240

Arg Gly Asn His Glu Lys Lys Lys Lys Lys 250
245 250

(2) INFORMATION FOR SEQ ID NO: 556:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 556:

20 Gly Leu Ala Gly Leu Cys Gly Gln Leu Ser Ser Pro Ala Leu Cys Val 15
1 5 10 15

Asn Arg Leu

25

(2) INFORMATION FOR SEQ ID NO: 557:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 217 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 557:

35 Met Ile Thr Glu Lys Trp Gly Leu Asn Met Glu Tyr Cys Arg Gly Gln 15
1 5 10 15

Ala Tyr Ile Xaa Ser Ser Gly Phe Ser Ser Lys Met Lys Val Val Ala 30
20 25 30

Ser Arg Leu Leu Glu Lys Tyr Pro Gln Ala Ile Thr Thr Leu Cys Ser 45
35 40 45

45 Ser Cys Ala Leu Asn Met Trp Leu Ala Lys Ser Val Pro Val Met Gly 60
50 55 60

Val Ser Val Ala Leu Gly Thr Ile Glu Glu Val Cys Ser Phe Phe His 80
65 70 75 80

50 Arg Ser Pro Gln Leu Leu Leu Glu Leu Asp Asn Val Ile Ser Val Leu 95
85 90 95

Phe Gln Asn Ser Lys Glu Arg Gly Lys Glu Leu Lys Glu Ile Cys His 110
100 105 110

Ser Gln Trp Thr Gly Arg His Asp Ala Phe Glu Ile Leu Val Glu Leu 125
115 120 125

60 Leu Gln Ala Leu Val Leu Cys Leu Asp Gly Ile Asn Ser Asp Thr Asn 15

130 135 140
Ile Arg Trp Asn Asn Tyr Ile Ala Gly Arg Ala Phe Val Leu Cys Ser
145 150 155 160
Ala Val Ser Asp Phe Asp Phe Ile Val Thr Ile Val Val Leu Lys Asn
165 170 175
Val Leu Ser Phe Thr Arg Ala Phe Gly Lys Asn Leu Gln Gly Gln Thr
180 185 190
Ser Asp Val Phe Phe Ala Ala Gly Ser Leu Thr Ala Val Leu His Ser
195 200 205
Leu Asn Glu Val Ile Gly Lys Tyr Xaa
210 215

(2) INFORMATION FOR SEQ ID NO: 558:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 558:

Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu Asn Glu
1 5 10 15
Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg Asn Thr
20 25 30
Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile Asn Phe
35 40 45
Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile Lys Leu
50 55 60
Tyr Thr Ser Lys Ser Glu Leu Pro Thr Asp Asn Ser Glu Thr Val Glu
65 70 75 80
Asn Thr

(2) INFORMATION FOR SEQ ID NO: 559:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 559:

Met Val Leu Ile Leu Leu Asn Leu Leu Leu Gly Gln Phe Ser Cys Met
1 5 10 15
Ser Pro Ala Ser His His Cys His Pro Leu Pro Thr Glu Met Pro Cys
20 25 30

Ser Ser Asp Trp Gly Phe Asp Ser His Thr Val Tyr Pro Ser Cys Val
35 40 45
Asp Ala Leu Leu Pro Lys Pro Ser Ala Asn Ser Phe Pro Asn Gly Ser
50 55 60
Cys His Cys Gln Gly Leu Tyr Asn Gln Gln Gln Asn Leu His Ala
65 70 75 80
Ala Glu Gly Pro Ala Ser Leu Arg Cys Asn Lys Tyr Val Ser Thr
85 90 95

(2) INFORMATION FOR SEQ ID NO: 560:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 560:

Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala Tyr Ala Ile Phe Phe Ile
1 5 10 15
Val Phe Thr Val Ile Gly Asp Ala Pro Gly Ala Val Leu Ser Cys Ala
20 25 30
Gly His Pro Cys Val Gly Phe Ala Ala Val Leu Val Ala Pro Leu Thr
35 40 45
Val Ala Val Ser Ser Xaa
50

(2) INFORMATION FOR SEQ ID NO: 561:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 108 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 561:

Met Glu Val Pro Pro Ala Pro Arg Ser Phe Leu Cys Arg Ala Leu
1 5 10 15
Cys Leu Phe Pro Arg Val Phe Ala Ala Glu Ala Val Thr Ala Asp Ser
20 25 30
Glu Val Leu Glu Glu Arg Gln Lys Arg Leu Pro Tyr Val Pro Glu Pro
35 40 45
Tyr Tyr Pro Glu Ser Gly Trp Asp Arg Leu Arg Glu Leu Phe Gly Lys
50 55 60
Asp Thr Val Asn Thr Ser Leu Asn Val Tyr Arg Asn Lys Asp Ala Leu
65 70 75 80

Ser His Phe Val Ile Ile Gly Ala Val Thr Gly Ser Leu Phe Arg Ile
85 90 95

5 Asn Val Gly Leu Arg Gly Trp Trp Leu Val Ala Xaa
100 105

10 (2) INFORMATION FOR SEQ ID NO: 562:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 562:

15 Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys Lys
1 5 10 15

20 Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala
20 25 30

25 Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Leu Leu Lys Met Arg Arg
35 40 45

Ala Pro
50

30 (2) INFORMATION FOR SEQ ID NO: 563:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 253 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 563:

40 Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu
1 5 10 15

Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu
20 25 30

45 Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp
35 40 45

Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr
50 55 60

50 Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu
65 70 75

55 Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu
85 90 95

Glu Ile Arg Gly Gly Glu Leu Lys Tyr Cys Asn Leu Val Arg
100 105 110

60 Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met

5 Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Arg
115 120 125
130 135 140

Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Leu Glu Asp Arg
145 150 155 160

10 Lys Tyr Glu Tyr Leu Met Thr Leu His Gly Val Val Asn Glu Ser Thr
165 170 175

Val Cys Leu Met Gly His Glu Arg Arg Gln Thr Leu Asn Leu Ile Thr
180 185 190

15 Met Leu Ala Ile Arg Val Leu Ala Asp Gln Asn Val Ile Pro Asn Val
195 200 205

Ala Asn Val Thr Cys Tyr Tyr Gln Pro Ala Pro Tyr Val Ala Asp Ala
210 215 220

20 Asn Phe Ser Asn Tyr Tyr Ile Ala Gln Val Gln Pro Val Phe Thr Cys
225 230 235 240

25 Gln Gln Gln Thr Tyr Ser Thr Trp Leu Pro Cys Asn Xaa
245 250

30 (2) INFORMATION FOR SEQ ID NO: 564:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 564:

Met Ser Phe Leu Met Trp Leu Met Ser Leu Ala Ile Thr Ser Gln Pro
1 5 10 15

40 Pro Met

45 (2) INFORMATION FOR SEQ ID NO: 565:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 565:

Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe
1 5 10 15

50 Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly
20 25 30

60 Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Phe Tyr Ser Ser
35 40 45

Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr
50 55 60
5 Gly Ala Ser Tyr His Ser Met Ser Ser Met Ala Arg Ala Ala Phe Phe
65 70 75 80

10

(2) INFORMATION FOR SEQ ID NO: 566:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 73 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 566:

His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val Gln Val Leu Ser
1 5 10 15

25

Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu Ala Pro Gly Arg
20 25 30

Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro Trp Phe Thr Ala
35 40 45

30

Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys Arg Gln Arg Arg
50 55 60

Gln Glu Arg Arg Gln Met Lys Arg Leu
65 70

35

(2) INFORMATION FOR SEQ ID NO: 567:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 263 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 567:

Met Asp Cys Pro Ala Leu Pro Pro Gly Trp Lys Lys Glu Glu Val Ile
1 5 10 15

Arg Lys Ser Gly Leu Ser Ala Gly Lys Ser Asp Val Tyr Tyr Phe Ser
20 25 30

Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln Leu Ala Arg Tyr Leu
35 40 45

Gly Asn Thr Val Asp Leu Ser Ser Phe Asp Phe Arg Thr Gly Lys Met
50 55 60

Met Pro Ser Lys Leu Gln Lys Asn Lys Gln Arg Leu Arg Asn Asp Pro
65 70 75 80

60

Leu Asn Gln Asn Lys Gly Lys Pro Asp Leu Asn Thr Thr Leu Pro Ile
85 90 95

Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val Thr Lys Val Thr Asn
100 105 110

His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln Arg Met Asn Glu Gln
115 120 125

Pro Arg Gln Leu Phe Trp Glu Lys Arg Leu Gln Gly Leu Ser Ala Ser
130 135 140

Asp Val Thr Glu Gln Ile Ile Lys Thr Met Glu Leu Pro Lys Gly Leu
145 150 155 160

Gln Gly Val Gly Pro Gly Ser Asn Asp Glu Thr Leu Leu Ser Ala Val
165 170 175

Ala Ser Ala Leu His Thr Ser Ser Ala Pro Ile Thr Gly Gln Val Ser
180 185 190

Ala Ala Val Glu Lys Asn Pro Ala Val Trp Leu Asn Thr Ser Gln Pro
195 200 205

Leu Cys Lys Ala Phe Ile Val Thr Asp Glu Asp Ile Arg Lys Gln Glu
210 215 220

Glu Arg Val Gln Gln Val Arg Lys Lys Leu Glu Glu Ala Leu Met Ala
225 230 235 240

Asp Ile Leu Ser Arg Ala Ala Asp Thr Glu Glu Met Asp Ile Glu Met
245 250 255

Asp Ser Gly Asp Glu Ala Xaa
260

35

(2) INFORMATION FOR SEQ ID NO: 568:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 70 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 568:

Met Met Arg Pro Phe Tyr Leu Leu Leu Pro Val Leu Cys Thr Gln Ala
1 5 10 15

Leu Arg Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Leu
20 25 30

Leu Phe Gly Leu Thr His Leu Asn Pro Ser Ala Lys Leu Leu Ser
35 40 45

Gln Met Lys Thr Ser Gly Asn Arg Lys Ser Glu Tyr Ser Lys Tyr Ala
50 55 60

Arg Asn Trp Lys Lys His
65 70

60

688

WO 98/39448

689

(2) INFORMATION FOR SEQ ID NO: 569:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 569:

Met Pro Val Thr Ser Lys Arg Thr Leu Phe Phe Pro Asp Pro Cys Ser
1 5 10 15

Tyr Asp Thr Pro Pro Asp Cys His Cys His Ser Phe Arg Ala Glu
20 25 30

Leu Leu

(2) INFORMATION FOR SEQ ID NO: 570:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 104 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 570:

Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu
1 5 10 15

Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala
20 25 30

Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu
35 40 45

His Ala Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ser
50 55 60

Lys Ala Pro Asn Ser Leu Gly Thr Thr Gly Leu Trp Ser Thr Val Tyr
65 70 75 80

Ile Phe Thr Glu Val Phe His Asn Phe Ser Asn Asn Ser Ile Phe Ser
85 90 95

Gly Lys Phe Leu Tyr Glu Val Asn
100

(2) INFORMATION FOR SEQ ID NO: 571:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 132 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 571:

Met Pro Val Leu Ser Leu Leu Cys Thr Leu Ile Val Ser Phe Gln Ser
1 5 10 15

Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu Leu His Ile Val Leu
1 5 10 15

Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala Trp Thr Leu Thr Asn
20 25 30

Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu His Ala Val Lys Gly
35 40 45

Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala Arg Leu Leu Thr His
50 55 60

Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr Ser Ser Arg Lys Phe
65 70 75 80

Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu Ala Ser Phe Tyr Thr
85 90 95

Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr Ala Ser Leu Leu Ser
100 105 110

Val Leu Ile Pro Lys Met Pro Gln Leu His Gly Val Arg Ile Phe Gly
115 120 125

Ile Asn Lys Tyr
130

(2) INFORMATION FOR SEQ ID NO: 572:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 572:

Met Asn Lys Trp Ile Cys Glu Met His Cys Tyr Leu Val Leu Leu Ser
1 5 10 15

Val Cys Ser Pro Ser Ala Leu Arg Val Arg His Thr Leu Ser Arg
20 25 30

(2) INFORMATION FOR SEQ ID NO: 573:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 573:

Met Pro Val Leu Ser Leu Leu Cys Thr Leu Ile Val Ser Phe Gln Ser
1 5 10 15

690

Ala Asp Ser Cys Glu Val Phe Leu Asn Cys Ser Leu
20 25

5

(2) INFORMATION FOR SEQ ID NO: 574:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 574:

Met Lys Val Ser Thr Met Leu Trp Phe Leu Cys Trp Glu Gln Ser His
1 5 10 15
Phe Leu Arg Glu Trp Glu Asp Leu Ser Thr Phe Leu Ile Leu Ile Gln
20 25 30

Met Glu Cys Gln Tyr Gly Asn Ser
35 40

25

(2) INFORMATION FOR SEQ ID NO: 575:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 575:

Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser
1 5 10 15
Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Xaa Leu
20 25 30

40

(2) INFORMATION FOR SEQ ID NO: 576:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 576:

Met Lys Arg Gly Cys Leu Gly Leu Phe Phe Ser Cys Cys Ser Ser
1 5 10 15
Ala Pro Thr Met Leu Leu Cys Asp Tyr Leu Asn Trp Phe
20 25

55

(2) INFORMATION FOR SEQ ID NO: 577:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 92 amino acids

60

691

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 577:

Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val Ala Ser Val
1 5 10 15

Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu Leu
20 25 30

Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu Lys
35 40 45

Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val Lys
50 55 60

Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Xaa Asn Arg Arg Arg
65 70 75 80

Xaa Val Arg Gly Leu Pro Ser Xaa Leu Thr Asn Ser
85 90

25

(2) INFORMATION FOR SEQ ID NO: 578:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 578:

Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg
1 5 10 15

Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg
20 25 30

Arg Val Met Val Asn Leu Asn Ile Leu Phe
35 40

45

(2) INFORMATION FOR SEQ ID NO: 579:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 579:

Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Phe Leu Val
1 5 10 15

Lys Leu Leu Phe Ile Pro Trp Leu Ala Ser Leu Leu Pro Leu Ser
20 25 30

Thr Phe Thr Ser Asp Phe Tyr Phe Met Glu Phe Gly Ile Glu Val Lys
35 40 45

Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Gln Lys Asn Thr Lys
50 55 60

Lys Phe Asn Lys Lys Lys
65 70

(2) INFORMATION FOR SEQ ID NO: 580:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 110 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 580:

Met Leu Arg Leu Leu Leu Val Ala Phe Ala Leu Val Val Val Leu
1 5 10 15

20 Phe His Val Leu Leu Ala Pro Ile Thr Ala Leu Phe His Thr His Phe
20 25 30

Tyr Asp Arg Leu Gln Asp Ala Gly Ser Arg Trp Pro Gln Leu Tyr Leu
35 40 45

25 Tyr Ser Arg Ala Asp Glu Val Val Leu Ala Arg Asp Ile Glu Arg Met
50 55 60

30 Val Glu Ala Arg Leu Ala Arg Arg Val Leu Ala Arg Ser Val Asp Phe
65 70 75 80

Val Ser Ser Ala His Val Ser His Leu Arg Asp Tyr Pro Thr Tyr Tyr
85 90 95

35 Thr Ser Leu Cys Val Asp Phe Met Arg Asn Cys Val Arg Cys
100 105 110

(2) INFORMATION FOR SEQ ID NO: 581:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 581:

Met Phe Lys Leu Gln Glu Cys Gly Lys Thr Thr Phe Leu Leu Ser Met
1 5 10 15

50 Ala Leu Tyr Phe Trp Trp Ile Val Gln Thr Thr Lys Gly Cys
20 25 30

(2) INFORMATION FOR SEQ ID NO: 582:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 582:

5 Met Glu Ser Asp Ala Leu Leu Thr Ile Phe Trp Ile Ile Ala Arg
1 5 10 15

Ser Ser Val Arg Ser Val Gly Lys Ser Ser Gln Arg Ser Phe Thr Thr
20 25 30

10 Ile Thr Gln Leu Arg Ser Thr His Thr Gly Pro Ser Arg Arg Ser Tyr
35 40 45

Leu Ile Trp Trp Asn Gly Gly Pro Lys Arg Thr Ile Ser Tyr Val Ser
50 55 60

15 Arg Arg Phe Arg Ser Phe Arg
65 70

(2) INFORMATION FOR SEQ ID NO: 583:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 583:

30 Val Gly Leu Phe Gln Pro Lys Thr Phe Gln Val Pro Val Thr Asp Leu
5 10 15

Tyr Ile Phe Ile Lys Ile Tyr Ser Glu Ile Gly Pro Ile Met His Val
20 25 30

35 Leu Cys Pro Gly Tyr Ser Gln Ser Pro Ser Thr Pro Pro Trp Thr
35 40 45

(2) INFORMATION FOR SEQ ID NO: 584:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 584:

Met Trp Phe Gly Ser Asp Arg Ser Asp Leu Arg Ile Gly Thr Ala Phe
1 5 10 15

50 Leu Phe Asp Leu Val Cys Asp Leu Cys Ile His Ala Trp Lys Pro Pro
20 25 30

55 Gly Leu Val Arg Phe Ser Phe
35

(2) INFORMATION FOR SEQ ID NO: 585:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 585:

Met Leu Asn Thr Ala Ser Leu Asn Leu Pro Trp Lys Val Gln Leu Phe
1 5 10 15

10 Ala His Ala

15 (2) INFORMATION FOR SEQ ID NO: 586:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 586:

Met Ser Ala Cys Leu Leu Phe Leu Ala Phe Ser Trp Lys Arg Lys
1 5 10 15

25 Gly Leu Trp Ser Gly Pro Gly
20

30 (2) INFORMATION FOR SEQ ID NO: 587:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:

Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser
1 5 10 15

Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val
20 25 30

Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro
35 40 45

Ser Ser Glu Leu Leu Gln Leu Leu Ser Val Gln Phe Val Trp Gln
50 55 60

50 Ala His Thr Val Ala
65

55 (2) INFORMATION FOR SEQ ID NO: 588:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

Met Gly Pro Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu
1 5 10 15

Leu Met Met Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu
20 25 30

Ser Thr Cys Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Asn Val Gly
35 40 45

Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys
50 55 60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 588:

Met Gly Pro Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu
1 5 10 15

Leu Met Met Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu
20 25 30

Ser Thr Cys Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu
35 40 45

Asn Val Gly Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr
50 55 60

Arg Lys Lys Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys
65 70 75

(2) INFORMATION FOR SEQ ID NO: 589:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 155 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 589:

Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe
1 5 10 15

Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val
20 25 30

Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro
35 40 45

Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala
50 55 60

Val Gly Phe Leu Glu Leu Leu Ala Gly Leu Leu Val Met Gly Pro
65 70 75 80

Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met
85 90 95

Gly Ala Ile Phe Thr Leu Ala Leu Lys Glu Ser Leu Ser Thr Cys
100 105 110

Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Asn Val Gly
115 120 125

Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys
130 135 140

Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa
145 150 155

696

(2) INFORMATION FOR SEQ ID NO: 590:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 590:

Met Pro Glu Thr Arg Leu Gly His Arg Gln Gln Phe Ala Val Phe His
 1 5 10 15

Leu Xaa Pro Val Pro Pro Cys Gly
 20

15

(2) INFORMATION FOR SEQ ID NO: 591:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 38 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 591:

Met Leu Thr Phe Leu Phe Ser Ala Cys Ala Thr Cys Leu Gly Lys Leu
 1 5 10 15

Ala Ser Pro Leu Ala Pro Val Gly Pro Gln Gln Arg Gly Xaa Pro Pro
 20 25 30

Gly Pro Pro Leu Leu Ser
 35

30

(2) INFORMATION FOR SEQ ID NO: 592:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 69 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 592:

Met Asp Pro Phe His Tyr Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu
 1 5 10 15

Val Phe Ala Val Val Leu Phe Ser Val Gly Ile Leu Leu Ile Leu Ser
 20 25 30

Arg Arg Cys Lys Cys Ser Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp
 35 40 45

Glu Glu Ala Gln Val Glu Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro
 50 55 60

Gln Lys Ala Glu Asn
 65

55

60

697

(2) INFORMATION FOR SEQ ID NO: 593:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 308 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 593:

Asn Leu Arg Val Arg Leu Gly Asp Val Ile Ser Ile Gln Pro Cys Pro
 1 5 10 15

Asp Val Lys Tyr Gly Lys Arg Ile His Val Leu Pro Ile Asp Asp Thr
 20 25 30

Val Glu Gly Ile Thr Gly Asn Leu Leu Phe Glu Val Tyr Leu Lys Pro Tyr
 35 40 45

Phe Leu Glu Ala Tyr Arg Pro Ile Arg Lys Gly Asp Ile Phe Leu Val
 50 55 60

Arg Gly Gly Met Arg Ala Val Glu Phe Lys Val Val Glu Thr Asp Pro
 65 70 75 80

Ser Pro Tyr Cys Ile Val Ala Pro Asp Thr Val Ile His Cys Glu Gly
 85 90 95

Glu Pro Ile Lys Arg Glu Asp Glu Glu Ser Leu Asn Glu Val Gly
 100 105 110

Tyr Asp Asp Ile Gly Gly Cys Arg Lys Gln Leu Ala Gln Ile Lys Glu
 115 120 125

Met Val Glu Leu Pro Leu Arg His Pro Ala Leu Phe Lys Ala Ile Gly
 130 135 140

Val Lys Pro Pro Arg Gly Ile Leu Leu Tyr Gly Pro Pro Gly Thr Gly
 145 150 155 160

Lys Thr Leu Ile Ala Arg Ala Val Ala Asn Glu Thr Gly Ala Phe Phe
 165 170 175

Phe Leu Ile Asn Gly Pro Glu Ile Met Ser Lys Leu Ala Gly Glu Ser
 180 185 190

Glu Ser Asn Leu Arg Lys Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro
 195 200 205

Ala Ile Ile Phe Ile Asp Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu
 210 215 220

Lys Thr His Gly Glu Val Glu Arg Arg Ile Val Ser Gln Leu Leu Thr
 225 230 235 240

Leu Met Asp Gly Leu Lys Gln Arg Ala His Val Ile Val Met Ala Ala
 245 250 255

Thr Asn Arg Pro Asn Ser Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg
 260 265 270

Phe Asp Arg Glu Val Asp Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu
 275 280 285

698

275 280 285
Glu Ile Leu Gln Ile His Thr Lys Asn Met Lys Leu Ala Asp Asp Val
290 300
Asp Leu Glu Gln
305

5

10

(2) INFORMATION FOR SEQ ID NO: 594:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 594:

Met Gln Ile Lys Leu Lys Ser Val Lys Thr Val Phe Ala Ile Thr
1 5 10 15
Leu Val Leu Phe Leu
20

25

(2) INFORMATION FOR SEQ ID NO: 595:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 595:

Met Phe Pro Lys Phe Cys Pro Ile Leu Ser Leu Val Asp Phe Ile Ser
1 5 10 15
His Arg Asp Lys Pro Glu Thr Glu
20

40

(2) INFORMATION FOR SEQ ID NO: 596:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 596:

Met Leu Ile Glu Cys Ala Trp Gln Leu Met Phe Leu Leu Lys Val
1 5 10 15
Glu Gln Leu Gly Ile Leu Asp Lys
20

55

(2) INFORMATION FOR SEQ ID NO: 597:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids

60

699

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 597:

Met
1

10

(2) INFORMATION FOR SEQ ID NO: 598:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 598:

Met Cys Ile Met Ser Ala Leu Val
1 5

20

(2) INFORMATION FOR SEQ ID NO: 599:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 599:

Met Phe Leu Val Trp Phe Trp Gly Leu Ile Ser Ala Leu Ser Asn
1 5 10 15
Val His Thr Pro Ser Arg Leu Pro Ala
20 25

35

(2) INFORMATION FOR SEQ ID NO: 600:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 600:

Met Xaa Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser
1 5 10 15
Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly
20 25

50

(2) INFORMATION FOR SEQ ID NO: 601:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids

60

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 601:

5 Met Trp Thr Arg Ser Ser Arg Cys Leu Leu Cys Ile Pro Gly Xaa
1 5 10 15

Ser Arg Arg Arg Ala Gly Ser Gly Met Lys Pro Arg Ser Trp Ser
20 25 30

10 Ala Trp Arg Pro Ser Gly Thr Gly Thr Ser Ser Gln Ser Ser
35 40 45

Thr Gln Ser Arg Thr Leu Ser Ala Thr Ala Ser Pro Ala
50 55 60

(2) INFORMATION FOR SEQ ID NO: 602:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 602:

25 Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu
1 5 10 15

30 Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr
20 25

35 (2) INFORMATION FOR SEQ ID NO: 603:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 603:

45 Met Pro Lys Gln Glu Leu Gly Ser Gly Val Gly Glu Leu Ala Lys
1 5 10 15

Asn Ser Lys Arg Gln His Trp Asn His Arg Trp Lys Lys Tyr Leu Lys
20 25 30

50 Leu Ile Arg Trp Glu Asp Gly Leu Leu Glu Gly Leu Leu Val
35 40 45

Leu Glu His Cys Ala Thr Met Ala Trp Asp Cys Leu Met Arg Leu Glu
50 55 60

55 Leu Leu Lys Arg Leu
65

60 (2) INFORMATION FOR SEQ ID NO: 604:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 604:

5 Lys Ile Val Tyr Ile Leu Gly Asn Pro Leu Lys Phe Asn Ser Arg Val
1 5 10 15

10 Ile His His Leu Val Leu Leu Gln
20

15 (2) INFORMATION FOR SEQ ID NO: 605:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 605:

20 Met Asn Leu His Gln Arg Arg Leu Leu Ile Gly His Leu Met Thr
1 5 10 15

25 Leu Val Lys Ala Ser Lys Ser Phe Thr Glu Ile Thr Ser Ser
20 25 30

30 Arg Lys Lys
35

35 (2) INFORMATION FOR SEQ ID NO: 606:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 606:

45 Leu Leu Gly Tyr Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr
1 5 10 15

Tyr Asn Ile His Leu His Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val
20 25 30

50 Gly Gly Leu Ser Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr
35 40 45

Val Gly Pro Thr Gln Arg Leu Leu Cys Gly Thr Leu Ala Ala Leu
50 55 60

55 His Met Leu Phe Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Xaa
65 70 75 80

Glu Gly Ile Leu Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln
85 90 95

60

Arg Val Pro Arg Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro
100 105 110

Thr Thr Val Leu Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln
115 120 125

Ser His
130

10

(2) INFORMATION FOR SEQ ID NO: 607:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 607:

Met Leu Val Ile Phe Leu Phe Thr Ser Leu Leu Lys Ile Pro Ser Ser
1 5 10 15

Val Pro Gly Leu Ile Asn Val
20

25

(2) INFORMATION FOR SEQ ID NO: 608:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 608:

Glu Leu Asp Tyr Ile Leu
1 5

35

(2) INFORMATION FOR SEQ ID NO: 609:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 232 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 609:

Met Ala Pro Pro Gly Tyr Gln Xaa Xaa Xaa Xaa Trp Leu Ala Cys
1 5 10 15

Pro Asp Arg Gly Glu Leu Ser Ser Arg Ser Pro Pro Cys Arg Leu Ala
20 25 30

55

Arg Trp Ala Glu Gly Asp Arg Glu Thr Arg Thr Cys Leu Leu Glu Leu
35 40 45

Ser Ala Gln Ser Trp Gly Gly Phe Arg Arg Ser Ser Ala Val Ser
50 55 60

60

Ala Gly Ser Pro Ser Arg Leu His Phe Leu Pro Gln Pro Leu Leu Leu
65 70 75 80

Arg Ser Ser Gly Ile Pro Ala Ala Thr Pro Trp Pro Gln Pro Ala
85 90 95

Gly Leu Pro Val Arg Pro Thr Arg Thr Gly Glu Asp Arg
100 105 110

Thr Leu Asp Ile Ser Ile Cys Thr Glu Val Leu Ala Gly Thr Glu Gln
115 120 125

Pro Pro Pro Arg Met Thr Ser Pro Ser Ser Ser Pro Val Phe Arg
130 135 140

Leu Glu Thr Leu Asp Gly Gly Gln Glu Asp Gly Ser Glu Ala Asp Arg
145 150 155 160

Gly Lys Leu Asp Phe Gly Ser Gly Leu Pro Pro Met Glu Ser Gln Phe
165 170 175

Gln Gly Glu Asp Arg Lys Phe Ala Pro Ser Asp Lys Ser Gln Pro Pro
180 185 190

Thr Thr Glu Arg Glu Gln Val Pro Val Ser Arg Ile Gln Thr Asp Leu
195 200 205

Thr Glu Ile Gly Ser Ser Met Arg Ser Pro Gly Val Ser Pro Arg Ile
210 215 220

Trp Leu Asp Phe Gln Ser Thr Xaa
225 230

35

(2) INFORMATION FOR SEQ ID NO: 610:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 610:

Met Val Leu Leu Leu Leu Leu Ala Tyr Val Leu Leu Thr Tyr Ile Leu
1 5 10 15

Leu Leu Asn Met Leu Ile Ala Leu Met Xaa Arg Asp Arg Gln Gln Cys
20 25 30

50

(2) INFORMATION FOR SEQ ID NO: 611:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 611:

Met Val Phe Glu Gly Phe Ser Ser Ala Phe Cys Leu Ser Ser Thr Ala
1 5 10 15
Pro Thr Ser His Pro
20

5

10

(2) INFORMATION FOR SEQ ID NO: 612:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 612:

Gly Lys Lys Asn Gln Leu Leu Val Ile
1 5

20

(2) INFORMATION FOR SEQ ID NO: 613:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 613:

Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val
1 5 10 15

30

35

Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys
20 25

40

(2) INFORMATION FOR SEQ ID NO: 614:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 614:

Met Ala Lys Arg Ser Pro Gly Gly Cys Gly Ser Gly Leu Ile Leu Leu
1 5 10 15

45

50

Cys Cys Gln Pro Cys Arg Pro Thr Ser Ser Ala Pro Met Arg
20 25 30

55

(2) INFORMATION FOR SEQ ID NO: 615:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 113 amino acids

(B) TYPE: amino acid

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 615:

Ile Thr Ile Ala Ile Gln Met Ile Cys Leu Val Asn Xaa Glu Leu Tyr
1 5 10 15

5

10

Pro Thr Phe Val Arg Asn Xaa Gly Val Met Val Cys Ser Ser Leu Cys
20 25 30

15

20

Asp Ile Gly Gly Ile Ile Thr Pro Phe Ile Val Phe Arg Leu Arg Glu
35 40 45

Val Trp Gln Ala Leu Pro Leu Ile Leu Phe Ala Val Leu Gly Leu Leu
50 55 60

25

30

Ala Ala Gly Val Thr Leu Leu Leu Pro Gln Thr Lys Gly Val Ala Leu
65 70 75 80

35

40

Pro Glu Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro
85 90 95

Lys Glu Asn Thr Ile Tyr Leu Lys Val Gln Thr Ser Glu Pro Ser Gly
100 105 110

45

50

Thr

(2) INFORMATION FOR SEQ ID NO: 616:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 616:

Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro Lys Glu
1 5 10 15

35

40

Asn Thr

(2) INFORMATION FOR SEQ ID NO: 617:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 617:

Pro Arg Val Arg Asn Ser Pro Glu Asp Leu Gly Leu Ser Leu Thr Gly
1 5 10 15

55

60

Asp Ser Cys Lys Leu
20

(2) INFORMATION FOR SEQ ID NO: 618:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 618:

10 Gln Ala Asp Asp Leu Gln Ala Thr Val Ala Ala Leu Cys Val Leu Arg
1 5 10 15

Gly Gly Gly Pro Trp Ala Gly Ser Trp Leu Ser Pro Lys Thr Pro Gly
20 25 30

15 Ala Met Gly Gly Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg
35 40 45

20 Lys Arg Leu Leu
50

(2) INFORMATION FOR SEQ ID NO: 619:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 619:

30 Gln Gln Glu Lys Ser Leu Ala Gly Trp Ala Leu Val Leu Ala Xaa
1 5 10 15

35 Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp Phe Gly Gly
20 25 30

Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr Leu Trp Leu
35 40 45

40 Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val Val Pro Gly
50 55 60

45 Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val Met Gly Val
65 70 75 80

Cys Cys Thr Ala Leu Leu Val Ala Val Val Ala Arg Lys Leu Glu Phe
85 90 95

50 Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp Ile Gln Tyr
100 105 110

Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln Glu Ala Trp
115 120 125

55 Met Phe Tyr Lys Lys His Thr Arg Arg Lys Glu Ser His Ala Ala Xaa
130 135 140

60 His Gln Arg Xaa Leu Leu Ala Ile Asn Ala Phe Arg Gln Val Arg
145 150 155 160

Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met Val Asp Ile
165 170 175

5 Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn Leu Ser Ser
180 185 190

Ser His Arg Ala Leu Glu Lys Lys Gln Ile Asp Thr Leu Ala Gly Lys Leu
195 200 205

10 Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly Pro Arg Gln Leu
210 215 220

Pro Glu Pro Ser Gln Ser Lys
225 230

(2) INFORMATION FOR SEQ ID NO: 620:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 620:

20 Tyr Gln Ala His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly
1 5 10 15

30 Thr Arg Gly Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser
20 25 30

Gly Ala Gly Lys
35

(2) INFORMATION FOR SEQ ID NO: 621:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 621:

45 Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu
1 5 10 15

Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser
20 25 30

Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala
35 40 45

55 His Lys Ala Lys Ser His Pro Glu Val
50 55

(2) INFORMATION FOR SEQ ID NO: 622:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 622:

Ile Thr Ser Thr Asp Ile Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln
1 5 10 15

Pro Ser Asp

(2) INFORMATION FOR SEQ ID NO: 623:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 623:

Asn Ser Thr Ser Gly Glu Cys Leu Leu Glu Ala Glu Gly Met Ser
1 5 10 15

Lys Ser Tyr

(2) INFORMATION FOR SEQ ID NO: 624:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 624:

Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys Ile Phe
1 5 10 15

Val Gly Ser Gly Ser Gly Thr Glu Gly Leu Val Met Asn Ser
20 25 30

Asp Ile Leu Gly Ala Thr Thr Glu Val Ile Glu Asp Ser Asp Ser
35 40 45

Ala Gly Pro

50

(2) INFORMATION FOR SEQ ID NO: 625:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 625:

Asp Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser
20 25 30

Ile Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala
1 5 10 15

His Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Leu Lys
20 25 30

Lys Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu
35 40 45

Gln Lys Gln Leu Leu Arg His Ala Lys His His Thr
50 55 60

(2) INFORMATION FOR SEQ ID NO: 626:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 626:

Asp Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala Arg Ala Phe Lys Ser
1 5 10 15

Ser His Asn Leu Ala Val His Arg Met Ile His Thr Gly Glu Lys
20 25 30

(2) INFORMATION FOR SEQ ID NO: 627:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 627:

Arg Ser Ser Arg Ser Lys Thr Gly Ser Leu Gln Leu Ile Cys Lys Ser
1 5 10 15

Glu Pro Asn Thr Asp Gln Leu Asp Tyr
20 25

(2) INFORMATION FOR SEQ ID NO: 628:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 183 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 628:

Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu
1 5 10 15

Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser
20 25 30

Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Asp Ser Val Val Ala
35 40 45

5 His Lys Ala Lys Ser His Pro Glu Val Xaa Ile Thr Ser Thr Asp Ile
50 55 60

10 Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln Pro Ser Asp Xaa Asn Ser
65 70 75 80

Thr Ser Gly Glu Cys Leu Leu Glu Ala Glu Gly Met Ser Lys Ser
85 90 95

15 Tyr Xaa Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys
100 105 110

Ile Phe Val Gly Ser Gly Ser Gly Gly Thr Glu Gly Leu Val Met
115 120 125

20 Asn Ser Asp Ile Leu Gly Ala Thr Thr Glu Val Leu Ile Glu Asp Ser
130 135 140

Asp Ser Ala Gly Pro Xaa Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala
145 150 155 160

25 Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met Ile His
165 170 175

30 Thr Gly Glu Lys His Tyr Xaa
180

(2) INFORMATION FOR SEQ ID NO: 629:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 629:

Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His
1 5 10 15

45 Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Leu Lys Lys
20 25 30

Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln
35 40 45

50 Lys Gln Leu Leu Arg His Ala Lys His His Thr Asp
50 55 60

(2) INFORMATION FOR SEQ ID NO: 630:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 630:

5 Pro Phe Lys Asp Asp Pro Arg Asp Glu Thr Tyr Lys Pro His Leu Glu
1 5 10 15

10 Arg Glu Thr Pro Lys Pro Arg Arg Lys Ser Gly
20 25

(2) INFORMATION FOR SEQ ID NO: 631:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 110 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 631:

20 Glu Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Leu
1 5 10 15

25 Met Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val
20 25 30

Leu Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu
35 40 45

30 Trp Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg
50 55 60

Met Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val
65 70 75 80

35 Gln Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu
85 90 95

40 Asn Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met
100 105 110

(2) INFORMATION FOR SEQ ID NO: 632:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 632:

50 Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg Pro Thr Gly Gln Gln
1 5 10 15

55 Leu Glu Ser Leu Gly Leu Leu Ala
20

(2) INFORMATION FOR SEQ ID NO: 633:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 633:

Val His Arg Glu Glu Ala Ser Cys Tyr Cys Gln Ala Glu Pro Ser Gly
1 5 10 15

10 Asp Leu

15 (2) INFORMATION FOR SEQ ID NO: 634:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 634:

Arg Pro Ala Leu Arg Gln Ala Gly Gly Thr Arg Glu Pro Arg Gln
1 5 10 15

25 Lys Arg Trp Ala Gly Leu
20

30 (2) INFORMATION FOR SEQ ID NO: 635:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 635:

Ala Val Asn Phe Arg Pro Gln Arg Ser Gln Ser Met
1 5 10

45 (2) INFORMATION FOR SEQ ID NO: 636:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 636:

Met Ile Thr Asp Val Gln Leu Ala Ile Phe Ala Asn Met Leu Gly Val
1 5 10 15

Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val Ala Val Asn Asn
20 25 30

Pro Lys Lys Gln Glu
35

60

(2) INFORMATION FOR SEQ ID NO: 637:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 342 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 637:

Glu Glu Met Ala Asp Ser Val Lys Thr Phe Leu Gln Asp Leu Ala Arg
1 5 10 15

Gly Ile Lys Asp Ser Ile Trp Gly Ile Cys Thr Ile Ser Lys Leu Asp
20 25 30

Ala Arg Ile Gln Gln Lys Arg Glu Gln Arg Arg Arg Ala Ser
35 40 45

Ser Val Leu Ala Gln Arg Arg Ala Gln Ser Ile Glu Arg Lys Gln Glu
50 55 60

Ser Glu Pro Arg Ile Val Ser Arg Ile Phe Gln Cys Cys Ala Trp Asn
65 70 75 80

Gly Gly Val Phe Trp Phe Ser Leu Leu Phe Tyr Arg Val Phe Ile
85 90 95

Pro Val Leu Gln Ser Val Thr Ala Arg Ile Ile Gly Asp Pro Ser Leu
100 105 110

His Gly Asp Val Trp Ser Trp Leu Glu Phe Phe Leu Thr Ser Ile Phe
115 120 125

Ser Ala Leu Trp Val Leu Pro Leu Phe Val Leu Ser Lys Val Val Asn
130 135 140

Ala Ile Trp Phe Gln Asp Ile Ala Asp Leu Ala Phe Glu Val Ser Gly
145 150 155 160

Arg Lys Pro His Pro Phe Pro Ser Val Ser Lys Ile Ile Ala Asp Met
165 170 175

Leu Phe Asn Leu Leu Leu Gln Ala Leu Phe Leu Ile Gln Gly Met Phe
180 185 190

Val Ser Leu Phe Pro Ile His Leu Val Gly Gln Leu Val Ser Leu Leu
195 200 205

His Met Ser Leu Leu Tyr Ser Leu Tyr Cys Phe Glu Tyr Arg Trp Phe
210 215 220

Asn Lys Gly Ile Glu Met His Gln Arg Leu Ser Asn Ile Glu Arg Asn
225 230 235 240

Trp Pro Tyr Tyr Phe Gly Phe Gly Leu Pro Leu Ala Phe Leu Thr Ala
245 250 255

Met Gln Ser Ser Tyr Ile Ile Ser Gly Cys Leu Phe Ser Ile Leu Phe
260 265 270

Pro Leu Phe Ile Ile Ser Ala Asn Glu Ala Lys Thr Pro Gly Lys Ala
275 280 285

5 Tyr Leu Phe Gln Leu Arg Leu Phe Ser Leu Val Val Phe Leu Ser Asn
290 295 300

Arg Leu Phe His Lys Thr Val Tyr Leu Gln Ser Ala Leu Ser Ser Ser
305 310 315 320

10 Thr Ser Ala Glu Lys Phe Pro Ser Pro His Pro Ser Pro Ala Lys Leu
325 330 335

15 Lys Ala Thr Ala Gly His
340

(2) INFORMATION FOR SEQ ID NO: 638:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 529 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 638:

Met Ala Lys Phe Met Thr Pro Val Ile Gln Asp Asn Pro Ser Gly Trp
1 5 10 15

30 Gly Pro Cys Ala Val Pro Glu Gln Phe Arg Asp Met Pro Tyr Gln Pro
20 25 30

Phe Ser Lys Gly Asp Arg Leu Gly Lys Val Ala Asp Trp Thr Gly Ala
35 40 45

35 Thr Tyr Gln Asp Lys Arg Tyr Thr Asn Lys Tyr Ser Ser Gln Phe Gly
50 55 60

Gly Gly Ser Gln Tyr Ala Tyr Phe His Glu Glu Asp Glu Ser Ser Phe
65 70 75 80

Gln Leu Val Asp Thr Ala Arg Thr Gln Lys Thr Ala Tyr Gln Arg Asn
85 90 95

45 Arg Met Arg Phe Ala Gln Arg Asn Leu Arg Arg Asp Lys Asp Arg Arg
100 105 110

Asn Met Leu Gln Phe Asn Leu Gln Ile Leu Pro Lys Ser Ala Lys Gln
115 120 125

50 Lys Glu Arg Glu Arg Ile Arg Leu Gln Lys Lys Phe Gln Lys Gln Phe
130 135 140

Gly Val Arg Gln Lys Trp Asp Gln Lys Ser Gln Lys Pro Arg Asp Ser
145 150 155 160

Ser Val Glu Val Arg Ser Asp Trp Glu Val Lys Glu Glu Met Asp Phe
165 170 175

60 Pro Gln Leu Met Lys Met Arg Tyr Leu Glu Val Ser Glu Pro Gln Asp

180 185 190
Ile Glu Cys Cys Gly Ala Leu Glu Tyr Tyr Asp Lys Ala Phe Asp Arg
195 200 205

5 Ile Thr Thr Arg Ser Glu Lys Pro Leu Arg Xaa Lys Arg Ile Phe
210 215 220

10 His Thr Val Thr Thr Asp Asp Pro Val Ile Arg Lys Leu Ala Lys
225 230 235 240

Thr Gln Gly Asn Val Phe Ala Thr Asp Ala Ile Leu Ala Thr Leu Met
245 250 255

15 Ser Cys Thr Arg Ser Val Tyr Ser Trp Asp Ile Val Val Gln Arg Val
260 265 270

Gly Ser Lys Leu Phe Phe Asp Lys Arg Asp Asn Ser Asp Phe Asp Leu
275 280 285

20 Leu Thr Val Ser Glu Thr Ala Asn Glu Pro Pro Gln Asp Glu Gly Asn
290 295 300

Ser Phe Asn Ser Pro Arg Asn Leu Ala Met Glu Ala Thr Tyr Ile Asn
305 310 315 320

His Asn Phe Ser Gln Gln Cys Leu Arg Met Gly Lys Glu Arg Tyr Asn
325 330 335

30 Phe Pro Asn Pro Asn Pro Phe Val Glu Asp Asp Met Asp Lys Asn Glu
340 345 350

Ile Ala Ser Val Ala Tyr Arg Tyr Arg Ser Gly Lys Leu Gly Asp Asp
355 360 365

35 Ile Asp Leu Ile Val Arg Cys Glu His Asp Gly Val Met Thr Gly Ala
370 375 380

Asn Gly Glu Val Ser Phe Ile Asn Ile Lys Thr Thr Leu Asn Glu Trp Asp
385 390 395 400

Ser Arg His Cys Asn Gly Val Asp Trp Arg Gln Lys Leu Asp Ser Gln
405 410 415

45 Arg Gly Ala Val Ile Ala Thr Glu Leu Lys Asn Asn Ser Tyr Lys Leu
420 425 430

Ala Arg Trp Thr Cys Cys Ala Leu Leu Ala Gly Ser Glu Tyr Leu Lys
435 440 445

50 Leu Gly Tyr Val Ser Arg Tyr His Val Lys Asp Ser Ser Arg His Val
450 455 460

Ile Leu Gly Thr Gln Gln Phe Lys Pro Asn Glu Phe Ala Ser Gln Ile
465 470 475 480

Asn Leu Ser Val Glu Asn Ala Trp Gly Ile Leu Arg Cys Val Ile Asp
485 490 495

60 Ile Cys Met Lys Leu Glu Glu Gly Lys Tyr Leu Ile Leu Lys Asp Pro

716

717

500 505 510
Asn Lys Gln Val Ile Arg Val Tyr Ser Leu Pro Asp Gly Thr Phe Ser
515 520 525
5 Ser
10
(2) INFORMATION FOR SEQ ID NO: 639:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 194 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 639:
Lys Lys Arg His Thr Asp Val Gln Phe Tyr Thr Glu Val Gly Glu Ile
1 5 10 15
Thr Thr Asp Leu Gly Lys His Gln His Met His Asp Arg Asp Asp Leu
20 25 30
Tyr Ala Glu Gln Met Glu Arg Glu Met Arg His Lys Leu Lys Thr Ala
35 40 45
Phe Lys Asn Phe Ile Glu Lys Val Glu Ala Leu Thr Lys Glu Glu Leu
50 55 60
Glu Phe Glu Val Pro Phe Arg Asp Leu Gly Phe Asn Gly Ala Pro Tyr
65 70 75 80
Arg Ser Thr Cys Leu Leu Gln Pro Thr Ser Ser Ala Leu Val Asn Ala
85 90 95
Thr Glu Trp Pro Phe Val Thr Leu Asp Glu Val Glu Leu Ile
100 105 110
His Phe Xaa Arg Val Gln Phe His Leu Lys Asn Phe Asp Met Val Ile
115 120 125
Val Tyr Lys Asp Tyr Ser Lys Lys Val Thr Met Ile Asn Ala Ile Pro
130 135 140
Val Ala Ser Leu Asp Pro Ile Lys Glu Trp Leu Asn Ser Cys Asp Leu
145 150 155 160
Lys Tyr Thr Glu Gly Val Gln Ser Leu Asn Trp Thr Lys Ile Met Lys
165 170 175
Thr Ile Val Asp Asp Pro Glu Gly Phe Phe Glu Gln Gly Gly Trp Ser
180 185 190
55 Phe Leu

60 (2) INFORMATION FOR SEQ ID NO: 640:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 amino acids
(B) TYPE: amino acid

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 640:
Arg Ser Gly Leu Gly Leu Gly Ile Thr Ile Ala Phe Leu Ala Thr Leu
1 5 10 15
Ile Thr Gln Phe Leu Val Tyr Asn Gly Val Tyr Gln Tyr Thr Ser Pro
20 25 30
Asp Phe Leu Tyr Ile Arg Ser Trp Leu Pro Cys Ile Phe Phe Ser Gly
35 40 45
Gly Val Thr Val Gly Asn Ile Gly Arg Gln Leu Ala Met Gly Val Pro
50 55 60
Glu Lys Pro His Ser Asp
65 70
(2) INFORMATION FOR SEQ ID NO: 641:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 101 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 641:
Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser Val Glu
1 5 10 15
Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Xaa Xaa Pro Xaa
20 25 30
Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser Phe Tyr
35 40 45
Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg Leu Phe
50 55 60
Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile Ser Asn
65 70 75 80
Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu Leu Asp
85 90 95
Thr Arg Arg Ser Gly
100
(2) INFORMATION FOR SEQ ID NO: 642:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 amino acids
(B) TYPE: amino acid

718

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 642:

Met Glu Ala Gln Gln Val Asn Glu Ala Glu Ser Ala Arg Glu Gln Leu 15
1 5 10
Gln Xaa Leu His Asp Gln Ile Ala Gly Gln Lys Ala Ser Lys Gln Glu 30
20 25
Leu Glu Thr Glu Leu Glu Arg Leu Lys Gln Glu Phe His Tyr Ile Glu 45
35
Glu Asp Leu Tyr Arg Thr Lys Asn Thr Leu Gln Ser Arg Ile Lys Asp 60
50 55
Arg Asp Glu Glu Ile Gln Lys Leu Arg Asn Gln Leu Thr Asn Lys Thr 80
65 70 75
Leu Ser Asn Ser Ser Gln Ser Glu Leu Glu Asn Arg Leu His Gln Leu 95
85 90
Thr Glu Thr Leu Ile Gln Lys Gln Thr Met Leu Glu Ser Leu Ser Thr 110
100 105
Glu Lys Asn Ser Leu Val Phe Gln Leu Glu Arg Leu Glu Gln Met 125
115 120
Asn Ser Ala Ser Gly Ser Ser Asn Gly Ser Ser Ile Asn Met Ser 140
130 135
Gly Ile Asp Asn Gly Glu Gly Thr Arg Leu Arg Asn Val Pro Val Leu 160
145 150 155
Phe Asn Asp Thr Glu Thr Asn Leu Ala Gly Met Tyr Gly Lys Val Arg 175
165 170
Lys Ala Ala Ser Ser Ile Asp Gln Phe Ser Ile Arg Leu Gly Ile Phe 190
180 185
Leu Arg Arg Tyr Pro Ile Ala Arg Val Phe Val Ile Ile Tyr Met Ala 205
195 200
Leu Leu His Leu Trp Val Met Ile Val Leu Leu Thr Tyr Thr Pro Glu 220
210 215
Met His His Asp Gln Pro Tyr Gly Lys 230
225 230

(2) INFORMATION FOR SEQ ID NO: 643:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 643:

Ile Arg His Glu Gln His Pro Asn Phe Ser Leu Glu Met His Ser Lys 15
1 5 10

719

Gly Ser Ser Leu Leu Phe Leu Pro Gln Leu Ile Leu Ile Leu Pro 30
20 25
5 Val Cys Ala His Leu His Glu Glu Leu Asn Cys 40
35
10 (2) INFORMATION FOR SEQ ID NO: 644:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 644:
Ser Phe Phe Ile Ser Glu Glu Lys Gly His Leu Leu Leu Gln Ala Glu 15
1 5 10
Arg His Pro Trp Val Ala Gly Ala Leu Val Gly Val Ser Gly Gly Leu 30
20 25
Thr Leu Thr Thr Cys Ser Gly Pro Thr Glu Lys Pro Ala Thr Lys Asn 45
35 40
Tyr Phe Leu Lys Arg Leu Leu Gln Glu Met His Ile Arg Ala Asn 60
50 55

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b1g)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97897
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b1g)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209043
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 119, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	September 4, 1997
Accession Number	209235
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97898
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bii)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209044
C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the International application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bii)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97899
C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the International application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 15, 1997
Accession Number	209045
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer	
Susan White POT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 130, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	April 28, 1997
Accession Number	209011
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 131, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97900
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97901
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page 131, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209046 C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209047 C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 137 . line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 22, 1997	Accession Number 209076
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 140 . line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit August 21, 1997	Accession Number 209215
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97904
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 134, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit July 3, 1997	Accession Number 209139
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 153, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209049
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 153, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97903
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit June 12, 1997 Accession Number 209119 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97902 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209048
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209050
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT <input checked="" type="checkbox"/> Further deposits are identified on an additional sheet	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 12, 1998
Accession Number	209627
C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the International application	
Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
 - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
 - (f) a polynucleotide which is a variant of SEQ ID NO:X;
 - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
 - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
 - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

10

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

15

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

20

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

25

9. A recombinant host cell produced by the method of claim 8.

30

10. The recombinant host cell of claim 9 comprising vector sequences.

35

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in

ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

10

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

20

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

25

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

30

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

35

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:
- 5 (a) contacting the polypeptide of claim 11 with a binding partner; and
(b) determining whether the binding partner effects an activity of the polypeptide.
21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.
- 10 22. A method of identifying an activity in a biological assay, wherein the method comprises:
- 15 (a) expressing SEQ ID NO:X in a cell;
(b) isolating the supernatant;
(c) detecting an activity in a biological assay; and
(d) identifying the protein in the supernatant having the activity.
23. The product produced by the method of claim 22.

KiSS-1, a Novel Human Malignant Melanoma Metastasis-Suppressor Gene

Jeong-Hyung Lee, Mary E. Miele, Deana J. Hicks, Karen K. Phillips, Jeffery M. Trent, Bernard E. Weissman, Danny R. Welch*

Background: Microcell-mediated transfer of chromosome 6 into human C8161 and MelJuSo melanoma cells suppresses their ability to metastasize by at least 95% without affecting their tumorigenicity. This observation demonstrates that the ability to metastasize is a phenotype distinct from tumor formation and suggests that tumorigenic cells acquire metastatic capability only after accumulating additional genetic defects. These results also imply that mutations of genes on chromosome 6 are among those late genetic changes responsible for metastatic potential. They further suggest that a melanoma metastasis-suppressor gene(s) is encoded on chromosome 6 or is regulated by genes on chromosome 6. **Purpose:** Our objective was to identify the gene(s) responsible for the suppression of metastasis in chromosome 6/melanoma cell hybrids. **Methods:** A modified subtractive hybridization technique was used to compare the expression of messenger RNAs (mRNAs), via an analysis of complementary DNAs (cDNAs), in metastatic cells (C8161 or MelJuSo) and nonmetastatic hybrid clones (neo6/C8161 or neo6/MelJuSo). **Results:** A novel cDNA, designated KiSS-1, was isolated from malignant melanoma cells that had been suppressed for metastatic potential by the introduction of human chromosome 6. Northern blot analyses comparing mRNAs from a panel of human melanoma cells revealed that KiSS-1 mRNA expression occurred only in nonmetastatic melanoma cells. Expression of this mRNA in normal heart, brain, liver, lung, and skeletal muscle was undetectable by northern blot analysis. Weak expression was found in the kidney and pancreas, but the highest expression was observed in the placenta. The KiSS-1 cDNA encodes a predominantly hydrophilic, 164 amino acid protein with a polyproline-rich domain indicative of an SH3 ligand (binds to the homology 3 domain of the oncoprotein Src) and a putative protein kinase C- α phosphorylation site. Transfection of a full-length KiSS-1 cDNA into C8161 melanoma cells suppressed metastasis in an expression-dependent manner. **Conclusions:** These data strongly suggest that KiSS-1 expression may suppress the metastatic potential of malignant melanoma cells. **Implications:** KiSS-1 may be a useful marker for distinguishing metastatic melanomas from

nonmetastatic melanomas. [J Natl Cancer Inst 1996;88:1731-7]

The spread of malignant tumor cells from a primary tumor to form metastases at distant sites is the most life-threatening complication of cancer and is responsible for the majority of deaths in affected individuals. Understanding the regulation of metastasis at the molecular level is required to devise new modalities of cancer therapy, particularly those that improve cure rates when treating metastatic cancer.

Metastasis is a multistep process involving complex interactions between tumor cells and host cells. To metastasize, tumor cells must invade from the primary tumor, dissociate from the tumor mass, and be transported to nearby or distant secondary sites. Single cells, homotypic clusters of cells, or heterotypic emboli then arrest at a distant site with the use of both organ-specific and nonspecific mechanisms, invade into the surrounding tissue, and respond to growth signals at the secondary site (1-4). A tumor cell must successfully accomplish each step in the pathway or metastases will not develop. Both positive and negative regulators exist for each step in the metastatic cascade (5-11), implicating the involvement of dozens of different genes. This provides the challenging task of identifying critical genes controlling the process for use as targets for therapeutic intervention.

To identify metastasis-controlling genes, we have used a functional assay of metastasis suppression of human tumor cells in athymic nude mice. We previously showed that the introduction of human chromosome 6 into human metastatic melanoma

*Affiliations of authors: J.-H. Lee, M. E. Miele, D. J. Hicks, D. R. Welch, Department of Experimental Pathology, The Jake Gittlen Cancer Research Institute, College of Medicine, The Pennsylvania State University, Hershey; K. K. Phillips, B.E. Weissman, Department of Pathology, Lineberger Cancer Center, University of North Carolina, Chapel Hill; J. M. Trent, Division of Intramural Research, National Center for Human Genome Research, Bethesda, MD.

Correspondence to: Danny R. Welch, Ph.D., Department of Experimental Pathology (C7810), College of Medicine, The Pennsylvania State University, 500 University Dr., Hershey, PA 17033-0850.

See "Notes" section following "References."

cell lines C8161 or MelJuSo by microcell-mediated chromosome transfer resulted in at least 95% suppression of metastasis without affecting tumorigenicity (12,13) or local invasiveness (14). Introduction of chromosome 1 also partially suppressed MelJuSo metastasis (13). The development of this panel of cell lines provided us with the necessary reagents to isolate genes specifically involved in the control of metastasis.

To identify the gene(s) responsible for metastasis suppression in malignant melanoma, we used a modified subtractive hybridization method (15-19). Candidate genes were identified by searching for a minimum 10-fold increase in messenger RNA (mRNA) expression in nonmetastatic chromosome 6-C8161 hybrid cells (neo6/C8161.1) compared with highly metastatic parental C8161 cells grown under similar conditions.

Materials and Methods

Cell Lines and Culture Conditions

C8161 is an amelanotic human melanoma cell line that metastasizes widely following intradermal, subcutaneous, or intravenous injection into athymic nude or severe combined immunodeficient mice (12,20). Subclones C8161c.9 and C8161c.8 were isolated from C8161 by limiting dilution and were chosen for these studies because they have the highest and lowest metastatic potentials, respectively (12). neo6/C8161.1, neo6/C8161.2, and neo6/C8161.3 (12) were derived from microcell-mediated chromosome transfer of a single copy of a neomycin-resistance gene-tagged human chromosome 6 with the use of the MCH262A1.D6 donor cell line (12,21). Four additional, independently derived neo6/C8161 hybrid cell clones (neo6/C8161.4, neo6/C8161.5b, neo6/C8161.6, and neo6/C8161.8) were subsequently developed. None of the neo6/C8161 hybrid cell clones were metastatic in athymic nude mice. neo6(del)(q21-q23) hybrid cell clones were prepared using a related chromosome 6 microcell donor variant A9/6q-cl.1.2 (22) and are metastatic in nude mice (Miele ME, Goldberg SF, Hyatt DL, Barbanti-Brodano G, Welch DR: manuscript in preparation).

Cell line nomenclature has been chosen to identify the origin and nature of each cell line as unambiguously as possible. Single-cell clones are preceded by a "c" (e.g., C8161c.9 is a single-cell clone), and pooled, uncloned populations are identified by a "p" (e.g., C8161-KiSS-1-P1 is an uncloned population No. 1 of C8161 cells transfected with KiSS-1). Microcell hybrids are identified by the tagged chromosome followed by a "c" [e.g., neo6/C8161.3 is single-cell clone 3 derived from a fusion with a neomycin-resistance gene-tagged human chromosome 6; neo6(del)(q21-q23)/C8161c.8 is single-cell clone 8 derived from a microcell fusion with a neotagged chromosome 6 containing deletion of the q21-q23 bands].

All cells were grown in Dulbecco's modified minimal Eagle's medium and Ham's F-12 medium (DME-F12) (Irvine Scientific Co., Santa Ana, CA) supplemented with 10% fetal bovine serum (cDME-F12). Neo-resistant cells, neo6/C8161, and neo6(del)(q21-q23)/C8161 hybrids were maintained in cDME-F12 containing 500 µg/mL geneticin (G-418; Life Technologies, Inc. [GIBCO BRL], Gaithersburg, MD). Nearly 70%-90% confluent cultures were passaged in Corning (Oneonta, NY) tissue culture dishes following detachment with a solution of calcium- and magnesium-free Dulbecco's phosphate-buffered saline containing 2 mM EDTA at split ratios of 1:10 to 1:20. All cultures were routinely tested and found to be negative for *Mycoplasma* spp. infection using a polymerase chain reaction (PCR)-based test kit (Pan Vera, Madison, WI).

Subtractive Hybridization

Subtractive hybridization was done as described (18,19,23) with minor modifications. Briefly, mRNAs were isolated from 80% to 90% confluent cells using the FastTrack mRNA isolation kit (Invitrogen Corp., San Diego, CA). A complementary DNA (cDNA) library was constructed from neo6/C8161.1 cells using a λZAP-cDNA Gigapack II Gold cloning kit (Stratagene, La Jolla, CA). Subtractive hybridization was carried out between the first-strand cDNA from nonmetastatic neo6/C8161.1 cells and mRNA from metastatic parental C8161 cells. cDNA was synthesized from 1 µg polyadenylic acid [poly(A)⁺]-enriched RNA from neo6/C8161.1 using oligo(dT) primers. Biotinylation of mRNA from

parental C8161 cells was performed using the Photobiotin-labeling system (Life Technologies, Inc.) according to manufacturer's instructions. To subtract, biotinylated mRNA (20 µg) was mixed with cDNA (1 µg) from neo6/C8161.1 and ethanol precipitated. The pellet was resuspended in diethylpyrocyanate-treated water (20.5 µL), then added to 2× hybridization buffer (80% formamide, 100 mM hydroxyethylpiperazine ethane sulfonic acid [pH 7.5], and 2% sodium dodecyl sulfate [SDS], 22.5 µL). This mixture was boiled for 2 minutes, chilled on ice for 5 minutes, and followed by the addition of NaCl (5 M, 2 µL). Hybridization was performed at 42 °C for 48 hours. Streptavidin (25 µg; Sigma Chemical Co., St. Louis, MO) was added to the hybridization mixture, and the mixture was incubated at room temperature for 5 minutes before extraction with phenol-chloroform-isoamyl alcohol (25:24:1). The organic phase was back-extracted twice with 50 µL streptavidin-binding buffer (100 mM Tris-HCl [pH 8.0], 1 mM EDTA, and 500 mM NaCl). The aqueous phases were then pooled. Unhybridized cDNA target was ethanol precipitated. This subtracted cDNA was used as a probe for the neo6/C8161.1 cDNA library screening following random primer labeling (Amersham Life Sciences, Inc., Arlington Heights, IL). Positive clones were isolated and used for evaluation of expression in northern blot analysis.

Screening Candidate cDNAs and Isolation of Full-Length KiSS-1

Candidate cDNAs were initially evaluated for differential expression by northern blot analysis using poly(A)⁺-enriched mRNAs from C8161 and neo6/C8161.1 cells. If the expression was greater in neo6/C8161.1 cells by at least 10-fold, the same cDNA was used to probe a more extensive blot containing RNAs from a panel of cell lines with differing metastatic potentials. If the pattern of expression correlated with metastatic potential, full-length cDNAs were isolated and further characterized.

Full-length KiSS-1 was obtained from a λZAP-cDNA library (7.5 × 10⁶ plaques) prepared from neo6/C8161.1 cells. After third-round screening, seven positive clones were isolated and identified as a same gene by restriction mapping and sequence analysis. Library screening was done using a 0.5-kilobase (kb) partial KiSS-1 probe obtained from the subtractive hybridization. The longest cDNA insert was sequenced on both strands by the dideoxy chain termination method with Sequenase version 2.0 (Amersham Life Sciences, Inc.).

Northern Blot Analysis

For northern blot analysis, poly(A)⁺-enriched mRNA was isolated from 80% to 90% confluent cells with the use of a FastTrack mRNA isolation kit (Invitrogen). Poly(A)⁺ mRNA (2.5 µg) was loaded and electrophoresed on 1% agarose gel containing 2.2 M formaldehyde at 78 V for 3.5 hours. RNA was transferred onto a nylon membrane using the Turboblott system (Schleicher & Schuell, Keene, NH) and fixed by UV cross-linking (Stratagene). Full-length KiSS-1 cDNA probe was radiolabeled using random priming (Amersham Life Sciences, Inc.). Prehybridization was done in 48% formamide, 4.8× sodium chloride/sodium citrate (SSC) (1× SSC = 0.15 M NaCl and 0.015 M sodium citrate), 20 mM Tris (pH 7.6), 1× Denhardt's solution (0.02% Ficoll, 0.02% bovine serum albumin, and 0.02% polyvinylpyrrolidone), 10% dextran sulfate, and 0.1% SDS. Hybridization was carried out in the same buffer plus herring sperm DNA (100 µg/mL) for 24 hours at 42 °C. The membrane was rinsed twice with 2× SSC containing 0.1% SDS for 15 minutes at 42 °C, followed by washing twice with 0.1× SSC containing 0.1% SDS for 15 minutes at 42 °C. The rinsed membrane was exposed to x-ray film (Du Pont-NEN, Boston, MA) for desired intensity.

Expression of KiSS-1 in normal human tissues was evaluated by use of a multiple-tissue RNA blot (2 µg of poly(A) RNA per lane) purchased from CLONTECH Laboratories, Inc. (Palo Alto, CA). The blot was hybridized by use of a full-length KiSS-1 probe as above. Expression of KiSS-1 in normal human melanocytes (CLONTECH Laboratories, Inc.) was determined by reverse transcription-polymerase chain reaction (RT-PCR). Detection of KiSS-1 message (1.0 kb) in placenta was possible following overnight exposure. KiSS-1 mRNA was detectable in the pancreas (0.8 kb) and kidney (0.9 kb) only after a 3-7-day exposure at -70 °C with two intensifying screens.

Transfections and Metastasis Assays

A full length of KiSS-1 cDNA was cloned into the pcDNA3 expression vector (Invitrogen Corp.) in which transcription is driven by human cytomegalovirus promoter. The resultant plasmid pcDNA3-KiSS-1 was transfected

into C8161 cells using Lipofectin (Life Technologies, Inc.). Individual transfectants were isolated and cloned following growth in cDME-F12 containing G-418 (500 µg/mL). Poly(A)⁺ mRNA was isolated from transfectant by use of the Micro-FastTrack mRNA isolation kit (Invitrogen Corp.), and northern blot analysis was done as described above. C8161-neo6 mix was isolated as a control, uncloned population of C8161 cells transfected with the pcDNA3 vector only.

Transfectants were evaluated for growth in vitro. Cells (2×10^4) were inoculated into 24-well tissue culture plates (Corning) in cDME-F12 medium. After 24, 48, 72, and 96 hours, cells were trypsinized and counted using a hemacytometer.

For spontaneous metastasis assays, cells (1×10^6) were injected intradermally into the dorsolateral flank of athymic nude mice (four to 12 mice per group). Female mice (3–4 weeks old) (Harlan Sprague-Dawley, Inc., Madison, WI) were used for these studies. Animals were maintained under the guidelines of the National Institutes of Health and The Pennsylvania State University College of Medicine. All protocols were approved by the Institutional Animal Care and Use Committee. Food and water were provided ad libitum. Tumor size was measured weekly by taking orthogonal measurements and were expressed as mean tumor diameter. Mean tumor diameter was calculated as described (12,13) by use of the following equation:

$$\sqrt{(\text{diameter}_x) \times (\text{diameter}_y)}$$

After the mean tumor diameter reached 1.5–2.0 cm, mice were killed, necropsied, and visible metastases were counted (20,24).

In the C8161 human melanoma model, the rank orders of spontaneous and experimental metastasis assays for multiple cell variants are equivalent (i.e., poorly metastatic clones in the spontaneous metastasis assay are poorly metastatic in experimental metastasis assays. Likewise, individual clones are highly metastatic in both assays (20)). For experimental metastasis assays, cells (2×10^5 to 3×10^5) suspended in ice-cold Hanks' balanced salt solution (0.2 mL) were injected into the lateral tail vein of 3- to 4-week-old female athymic nude mice. After 4 weeks, the mice were killed by cervical dislocation, and metastatic lesions were scored as described (20,24). Quantification of metastasis formation was identical, regardless of the metastasis assay used. Briefly, the lungs from each mouse were removed, fixed in a mixture of formalin and Bouin's fixative (5:1 vol/vol), and examined under a dissecting stereomicroscope (24). Unless otherwise noted, all other tissues were examined and found to be free of metastases.

Statistical Analysis

The number of lung metastases was compared in KiSS-1 transfectants and parental C8161 cells. For experimental metastasis assays, one-way analysis of variance (ANOVA) followed by Tukey's Honestly Significant Difference post-test was used. For spontaneous metastasis assays, a Kruskal-Wallis ANOVA of ranks procedure was used. Calculations were performed using SigmaStat statistical analysis software (Jandel Scientific, San Rafael, CA). Statistical significance was defined as $P \leq 0.05$ using two-sided tests.

Results and Discussion

Several candidate metastasis-suppressor cDNAs were identified by subtractive hybridization and differential display comparing C8161 and neo6/C8161.1 cells (25). One clone, designated KiSS-1, was expressed only in nonmetastatic neo6/C8161.1 cells. The cDNA designation combines interim laboratory nomenclature for putative Suppressor Sequences with acknowledgment of the gene's discovery in Hershey. KiSS-1 cDNA sequence was submitted to GenBank as a novel gene with an accession number of U43527.

Lack of Expression of KiSS-1 in Metastatic Melanoma Cells

The expression pattern of KiSS-1 was confirmed by northern blot analysis using an extensive panel of cell lines derived from the C8161 melanoma with widely ranging metastatic potentials. KiSS-1 mRNA expression could not be detected in any meta-

static melanoma cell line (Fig. 1, A). Expression was not detectable, even in overexposed blots or by RT-PCR, in parental C8161 or two subclones, C8161cl.9 and C8161cl.8, representing the highest and lowest metastatic potential among clones, respectively (12). KiSS-1 mRNA expression was also undetectable in C8161cl.9 microcell hybrids containing human chromosome 6 bearing a complex deletion within the region 6q21→q23 that remained metastatic in nude mice (Miele ME, Goldberg SF, Hyatt DL, Barbanti-Brodano G, Welch DR: manuscript in preparation).

Seven independently prepared hybrids that contained an intact copy of human chromosome 6 in C8161 expressed high levels of a 1.0-kb transcript of KiSS-1 mRNA. Equal loading of lanes was confirmed by measuring glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression. All neo6/C8161 hybrid clones failed to metastasize following intravenous (experimental metastasis assay), subcutaneous, or intradermal (spontaneous metastasis assay) injection into 3- to 4-week-old female athymic nude mice. Therefore, KiSS-1 mRNA expression appeared to be a qualitative marker—i.e., exclusive to nonmetastatic C8161 cell populations.

Expression of KiSS-1 mRNA was evaluated by northern blotting of other, unrelated human metastatic melanoma cell lines (Fig. 1, B). To test for the possibility that truncated forms of KiSS-1 may have existed, a full-length cDNA probe was used. We could not detect the 1.0-kb KiSS-1 transcript or bands of other sizes in any cells capable of metastasizing in athymic nude mice.

The expression of KiSS-1 mRNA in the human heart, brain, placenta, liver, lung, skeletal muscle, kidney, or pancreas was also examined by northern blot analyses (Fig. 1, C). Abundant KiSS-1 transcript (1.0 kb) was found in the placenta, with very weak expression in the kidney (detectable only after exposure for ≥ 3 days using two intensifying screens). Upon longer exposure, different transcript sizes of KiSS-1 were detected in the pancreas (0.8 kb) and the kidney (0.9 kb), suggesting that alternative splicing may take place in different tissues. Expression of KiSS-1 could also be detected in normal human melanocytes by RT-PCR (data not shown). Expression in melanocytes suggests that KiSS-1 functions in normal precursor cells and may therefore be useful in staging melanoma progression. The precise role of KiSS-1 in melanocyte function has not yet been determined.

KiSS-1 cDNA has a single open reading frame that encodes a protein of 164 amino acids with a predicted molecular mass of 18 kd (Fig. 2). The initiation codon and surrounding nucleotides fit the Kozak consensus, and there is a consensus polyadenylation site downstream of the termination codon. In vitro transcription and translation of a full-length KiSS-1 cDNA resulted in a single band at 18 kd (data not shown). A search of the GenBank and European Molecular Biology Laboratory databases revealed no significant homology with known genes. A similar lack of homology was seen for protein sequences in the SwissBank database.

Metastasis Suppressed in C8161 KiSS-1 Transfectants

To demonstrate a functional basis for the correlation of KiSS-1 expression and the lack of metastatic potential, full-length

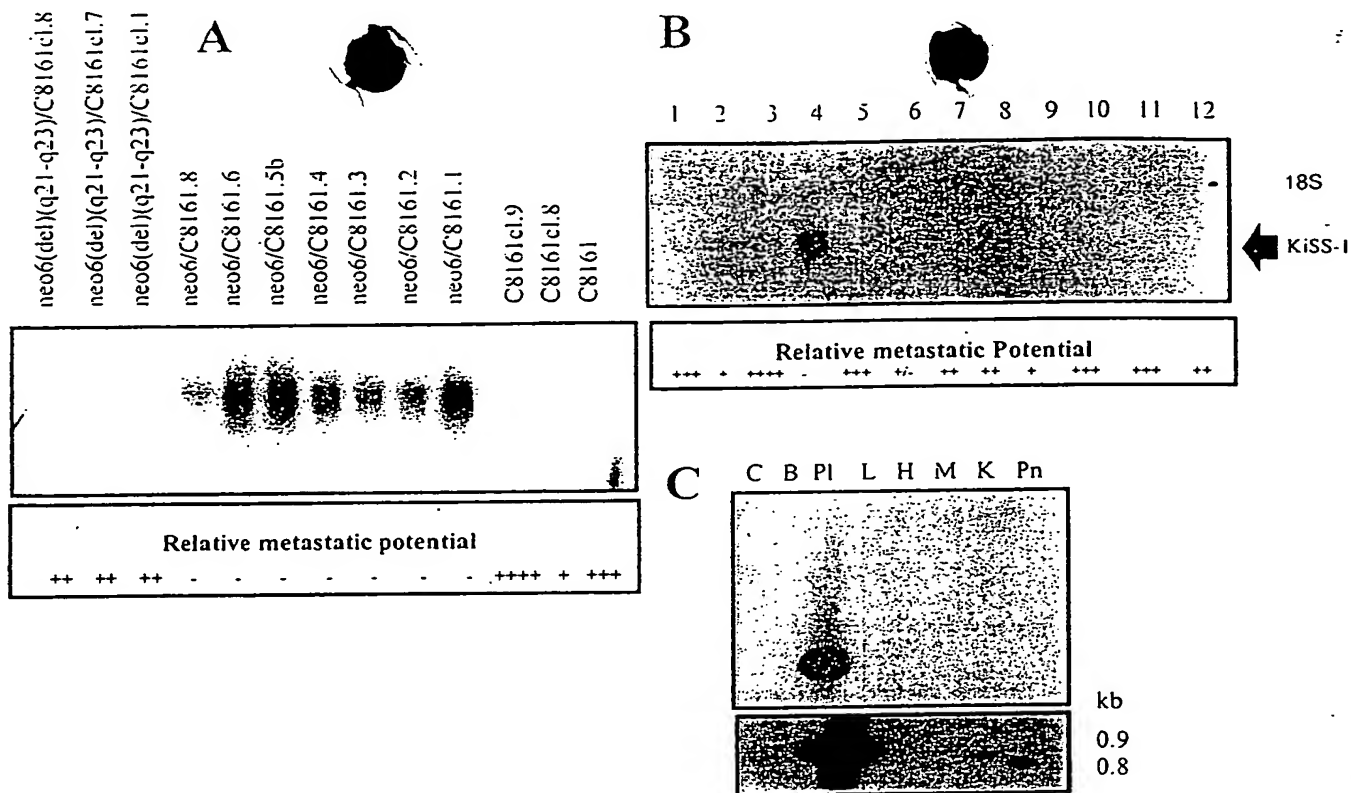


Fig. 1. Northern blot analyses of KiSS-1. **A**) Expression of KiSS-1 messenger RNA (mRNA) in cell clones derived from C8161, neo6/C8161, and neo6(del)(q21-q23)/C8161 cell clones. neo6/C8161 hybrids were prepared using MCH262A1.D6 as a microcell chromosome donor (27), and neo6(del)(q21-q23) hybrids were prepared using a related variant, A9/6q Cl.1.2 (22), as a microcell chromosome donor (Miele ME, Goldberg SF, Hyatt DL, Barbanti-Brodano G, Welch DR; manuscript in preparation). Relative metastatic potentials are based on both experimental (intravenous inoculation into lateral tail vein) and spontaneous (intradermal inoculation into dorsolateral flank) metastasis assays. Poly(A)⁺ mRNA (2.5 µg) was electrophoresed on denatured agarose gel, transferred onto a nylon membrane and fixed and probed with full-length KiSS-1 cDNA. KiSS-1 was undetectable in any cell line possessing metastatic potential, whether low or high. **B**) Expression of KiSS-1 mRNA in human metastatic melanoma cell lines. Lane 1 = C8161; 2 = C8161cl.8; 3 = C8161cl.9; 4 = non-

metastatic hybrid cell clone neo6/C8161.1; 5 = MelJuSo; 6 = A375P; 7 = A375M; 8 = MeWo; 9 = MeWo-3S5; 10 = MeWo-70W; 11 = M24met; and 12 = OM431. Poly(A)⁺ RNA was isolated from each cell culture, and northern blot analysis was performed as above. All cell lines having metastatic potential did not express KiSS-1. **C**) Expression of KiSS-1 mRNA in normal human tissues. The multiple-tissue RNA blot [2 µg of poly(A)⁺ RNA per lane] was purchased from CLONTECH Laboratories, Inc., Palo Alto, CA. The blot was hybridized by use of a full-length KiSS-1 probe as above. Detection of KiSS-1 message (1.0 kilobase [kb]) in the placenta was possible following overnight exposure. KiSS-1 mRNA was detectable in the pancreas (0.8 kb) and in the kidney (0.9 kb) only after a 3-7-day exposure at -70 °C with two intensifying screens. Besides differential organ expression, these results suggest that alternative splice variants of KiSS-1 may exist in different cells.

cDNA of KiSS-1 was subcloned into the pcDNA3 constitutive expression vector and transfected into C8161. The vector alone was transfected as a negative control. Several clones were randomly selected and analyzed for expression of KiSS-1 by RNA blotting (Fig. 3). The KiSS-1 transcript in transfectants is larger (1.3 kb) because of additional sequences in the expression vector. The in vitro growth rates of the selected clones were not significantly ($P > .05$) different compared with parental metastatic C8161 cell clones (Fig. 4). Clones expressing differing levels of KiSS-1 transcript were tested for metastatic ability in athymic nude mice by use of the experimental and spontaneous metastasis assays.

In the spontaneous metastasis assay, which measures the ability of cells injected intradermally or subcutaneously into the dorsolateral flank to metastasize to distant sites (20,24), KiSS-1 transfectant clones 2, 3, and 9 were less able to colonize lung or

regional lymph nodes than concomitantly injected C8161 cells (Table 1). Differences in metastatic potential could not be explained by slower in vivo primary tumor growth rate, since several more metastatic subclones of C8161 grow more slowly than KiSS-1 transfectants (12). Parental C8161 cells yielded an average of 50 lung metastases per mouse, and every mouse had regional lymph node metastases. In contrast, transfectant clone KiSS-1cl.2, which highly expressed KiSS-1 RNA, produced an average of only one metastasis per mouse. None of the mice had more than three metastases, and only two had lymph node metastases. KiSS-1 cl.3 produced a mean of three metastases per mouse (all mice yielded less than six metastases) and slightly higher incidences of lymph node metastases. Similar results were obtained following direct inoculation into the lateral tail veins of athymic nude mice (experimental metastasis assay, Table 2).

Predicted KiSS-1 Protein Possibly an SH3 Ligand or Phosphorylated by PKC- α

The predicted protein sequence encoded by KiSS-1 cDNA was analyzed, and the following homologies were found: a putative PKC- α phosphorylation site (single-letter amino acid code, SMR, corresponding to amino acid positions 154-156), and a proline-rich region with five overlapping minimal SH3 binding domains (binding to the homology 3 domain of the oncoprotein Src) (26,27) (PXXP motif, Fig. 2, B, corresponding to amino acids 85-97, where P is proline and X can be any amino acid). This suggests that KiSS-1 may be a ligand for a protein possessing an SH3 domain and/or may be a substrate for protein kinase C- α phosphorylation. Additionally, there are four cysteines, leaving the possibility for intramolecular or intermolecular disulfide linkages. The N-terminal 70 amino acids contain seven serines and a tyrosine in position 60 as possible phosphorylation sites. There are also five threonines in the KiSS-1 molecule. There are no apparent transmembrane, glycosylation, nuclear translocation, or domains indicative of KiSS-1 being a kinase. Although predominantly hydrophilic, it is unlikely that KiSS-1 is a secreted protein on the basis of the rules of von Heijne (28).

The PXXP motif is a constant feature of all SH3 ligands. The presence of five PXXP sequences in a 13 amino acid stretch is further evidence that KiSS-1 is an SH3 ligand (26,27,29,30). Rules governing specificity of SH3 binding are currently being elucidated (26,27,31-35), but for the most part, they remain unknown. However, the putative KiSS-1 SH3 binding domain shares many features of the RLP-type (class I) ligand orientation.

SH3 domains are 50-70 amino acid modules present in a variety of intracellular proteins that mediate protein-protein in-

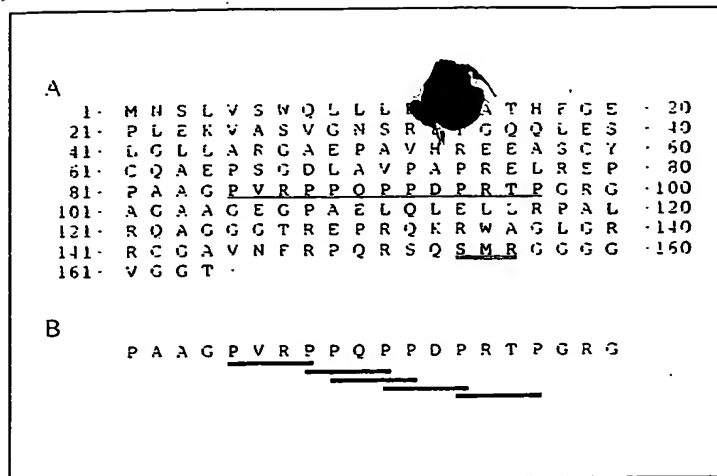


Fig. 2. A) Predicted protein sequence of KiSS-1. The nucleotide sequence has been submitted to GenBank (accession number U43527). Abbreviations for amino acid residues are: A = Ala; C = Cys; D = Asp; E = Glu; F = Phe; G = Gly; H = His; I = Ile; K = Lys; L = Leu; M = Met; N = Asn; P = Pro; Q = Gln; R = Arg; S = Ser; T = Thr; V = Val; W = Trp; and Y = Tyr. Full-length KiSS-1 was obtained from the complementary DNA (cDNA) library screening using a neo6/C8161.1 λ ZAP-cDNA library (7.5×10^6 plaques). After third-round screening, seven positive clones were isolated and identified as the same gene by restriction mapping and sequence analysis. The longest cDNA insert was sequenced on both strands by the dideoxy chain termination method with Sequenase version 2.0 (Amersham Life Sciences, Inc., Arlington Heights, IL). KiSS-1 has a polyproline-rich domain (underlined) and a putative protein kinase C- α phosphorylation site (double underlined). B) Amino acids 81-100 showing PXXP consensus minimal SH3 binding motif (underlined).

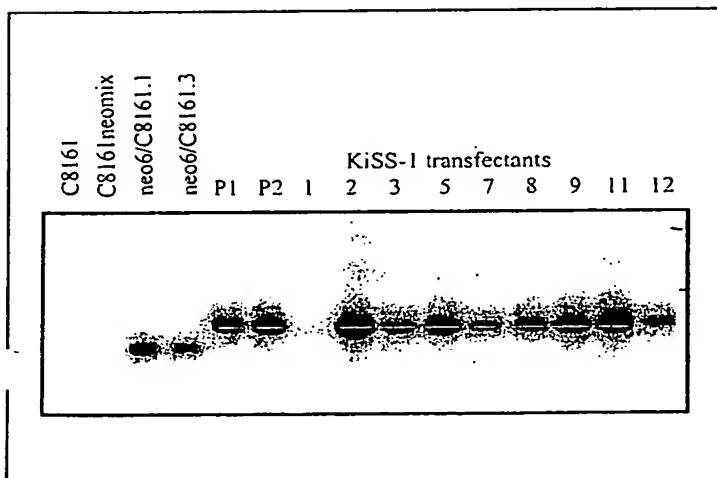


Fig. 3. Expression of KiSS-1 messenger RNA (mRNA) in KiSS-1 transfectants. A full length of KiSS-1 complementary DNA was cloned into pcDNA3 (Invitrogen Corp., San Diego, CA) in which transcription is driven by human cytomegalovirus promoter. The resultant plasmid pcDNA3-KiSS-1 was transfected into C8161 cells using Lipofectin (Life Technologies, Inc., [GIBCO-BRL] Gaithersburg, MD). Individual transfectants were isolated and cloned following growth in the neomycin analog G-418 (600 μ g/mL). Polyadenylic acid [Poly(A)] mRNA was isolated from each transfectant using Micro-Fast-Track mRNA isolation-kit (Invitrogen Corp.), and northern blot analysis was done as described above. C8161neo6 is an uncultured population of C8161 cells transfected with pcDNA3 vector only. Two different-sized transcripts are observed. KiSS-1 (1.0 kilobase [kb]) is smaller than exogenous KiSS-1 (1.3 kb) because of vector sequences in the latter. Lanes labeled P1 and P2 are uncultured populations of C8161 cells transfected with pcDNA3-KiSS-1 construct.

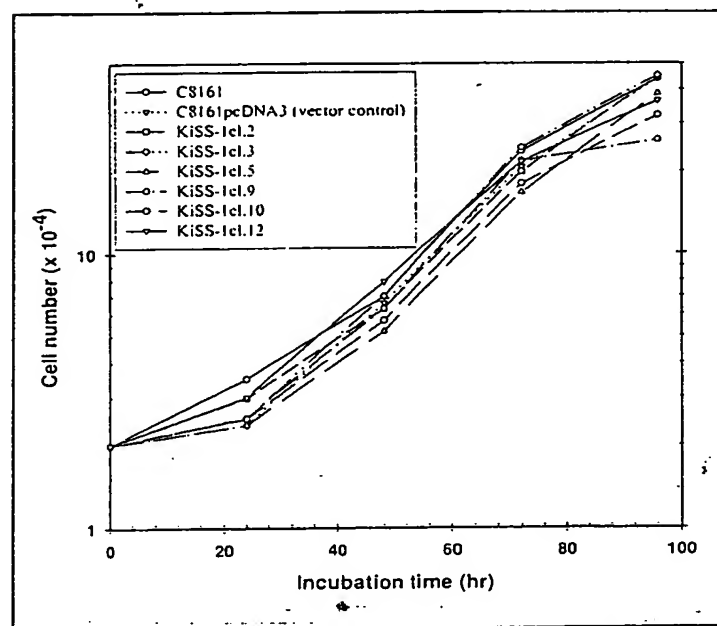


Fig. 4. Growth of KiSS-1 transfectants in vitro. Cells (2×10^4) were inoculated into a 24-well tissue culture plate in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F-12 medium supplemented with 10% fetal bovine serum. After 24, 48, 72, and 96 hours, cells were trypsinized and counted with a hemacytometer. Data are shown as mean \pm standard error.

Table 1. Suppression of spontaneous metastasis of C8161 human melanoma cells by KiSS-1*

Cell line	Relative KiSS-1 expression	No. of mice with metastases/total No. of mice	Lung		Extrapulmonary
			Mean \pm SEM	Median (range)	
C8161	-	16/16	50 \pm 25	9.5 (2 to >200)	8/8
KiSS-1cl.2	+++	5/7	1.1 \pm 0.4	1 (0 to 3)	2/8
KiSS-1cl.3	+	6/7	2.7 \pm 0.7	3 (0 to 6)	4/8
KiSS-1cl.9	++	0/7	0	0	2/7

*Cells (1×10^6) were injected intradermally into the dorsolateral flank of 3- to 4-week-old female athymic nude mice (seven to eight mice per group, Harlan Sprague-Dawley). When the mean tumor diameter (square root of the product of orthogonal measurements) reached 1.5-2.0 cm, mice were killed. All organs were examined for the presence of metastases, and many were confirmed by microscopic examination of hematoxylin-eosin-stained paraffin-embedded sections (4-6 μ m). Macroscopic lung metastases were quantified after being stained with a mixture of neutral-buffered formalin and Bouin's fixative (5:1) and were counted with the aid of a dissecting microscope as previously described (20,24). SEM = standard error of the mean.

interactions that are important for intracellular signaling and cytoskeletal organization (36-43). Many of these interactions involving SH3 domains have been directly or indirectly associated with various steps in the metastatic cascade. Therefore, KiSS-1 could suppress metastasis by regulating key signaling pathways important to one or more of these steps. KiSS-1 transfectants did not display significantly ($P > .05$) influenced adhesion to the extracellular matrix components type-I collagen or fibronectin nor was adhesion to a complex basement membrane-like mixture, Matrigel, affected (data not shown).

KiSS-1 maps to chromosome 1q32-q41 by fluorescence in situ hybridization (data not shown). Its mapping to another location within the genome compels the hypothesis that KiSS-1 is regulated by the gene(s) on chromosome 6. Our data also suggest that KiSS-1 is an important downstream effector of a gene(s) encoded on chromosome 6 and are consistent with metastasis suppression following introduction of chromosome 1 into MelJuSo melanoma cells (13). This hypothesis can be further refined because of data failing to detect KiSS-1 mRNA in metastatic neo6(del)(q21-q23)/C8161 hybrids. The latter can be explained by: 1) KiSS-1 mutations, 2) KiSS-1 deletions in subclones of C8161, or 3) mutation or deletion of regulatory

gene(s) mapping to 6q21-q23. These hypotheses will require additional studies. The latter explanation is consistent with high-frequency loss of heterozygosity observed in late stage melanomas (44-46).

In summary, we cloned a new human melanoma metastasis-suppressor gene by subtractive hybridization comparing metastatic C8161 cells and nonmetastatic neo6/C8161.1 cells. Loss of expression correlates with metastatic potential in human melanoma cells, and its expression results in significant suppression of melanoma metastasis in athymic nude mice. It will be necessary to further analyze the KiSS-1 gene in melanoma progression. The predicted KiSS-1 protein has a proline-rich region with homology to SH3-binding domain that could predict a mechanism for melanoma metastasis suppression. KiSS-1 may function in cell signaling and/or cytoskeletal organization on the basis of the usual role of SH3 domains in these processes; however, there is no evidence directly supporting these hypotheses. Taken together, these data strongly support a critical role for KiSS-1 in the regulation of cancer metastasis in human malignant melanoma. Additional experiments will be required to determine a role for KiSS-1 in other cancers or whether KiSS-1 may be a useful marker for staging melanomas in a clinical setting.

Table 2. Suppression of experimental metastasis of C8161 human melanoma cells by KiSS-1*

Cell line	Relative KiSS-1 expression	No. of mice with lung metastases/total No. of mice	Lung metastases		P
			Mean \pm SEM	Median (range)	
C8161	-	12/12	152 \pm 25	>200 (2 to >200)	
C8161cl.9	-	4/4	200 \pm 41	>200 (98 to >200)	NSD
C8161-neomix	-	4/4	160 \pm 26	174 (93 to >200)	NSD
KiSS-1cl.12	+	8/8	81 \pm 31	40 (1 to >200)	$P < .05$
KiSS-1cl.3	+	16/16	82 \pm 49	101 (2 to >200)	$P < .05$
KiSS-1cl.9	++	15/16	21 \pm 13	24.5 (0 to 80)	$P < .001$
KiSS-1cl.10	+++	4/12	1 \pm 1	0 (0 to 4)	$P < .0001$
KiSS-1cl.2	+++	9/16	0.9 \pm 0.4	0.5 (0 to 3)	$P < .001$

*Cells (3×10^5) were injected intravenously into the lateral tail vein of 3- to 4-week-old female athymic nude mice (seven to eight mice per group, Harlan Sprague-Dawley). Mice were killed 26 days after inoculation. All organs were examined for the presence of metastases, and many were confirmed by microscopic examination of hematoxylin-eosin-stained paraffin-embedded sections (4-6 μ m). Macroscopic lung metastases were quantified with the aid of a dissecting microscope after staining with a mixture of neutral-buffered formalin and Bouin's fixative (5:1) as described (20,24). P values were calculated by one-way analysis of variance using Tukey's Honestly Significant Difference post-test. Values shown compare each group with parental C8161. SEM = standard error of the mean; NSD = no significant difference.

References

- (1) Liotta LA, Stetler-Stevenson WG. Tumor progression and metastasis: an imbalance of positive and negative regulation. *Cancer Res* 1991;51(18 Suppl):5054s-9s.
- (2) Liotta LA, Steeg PS, Stetler-Stevenson WG. Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation. *Cell* 1991; 64:327-36.
- (3) Fidler IJ, Ellis LM. The implications of angiogenesis for the biology and therapy of cancer metastasis [see comment citation in Medline]. *Cell* 1994; 79:185-8.
- (4) Nicolson GL. Cancer progression and growth: relationship of paracrine and autocrine growth mechanisms to organ preference of metastasis. *Exp Cell Res* 1993;204:171-80.
- (5) Dear TN, Kefford RF. Molecular oncogenetics of metastasis. *Mol Aspects Med* 1990;11:243-324.
- (6) Wright JA, Egan SE, Greenberg AH. Genetic regulation of metastatic progression. *Anticancer Res* 1990;10:1247-55.
- (7) MacDougall JR, Matrisian LM. Contributions of tumor and stromal matrix metalloproteinases to tumor progression, invasion and metastasis. *Cancer Metastasis Rev* 1995;14:351-62.
- (8) Fodstad O, Kjonniksen I. Microenvironment revisited: time for reappraisal of some prevailing concepts of cancer metastasis. *J Cell Biochem* 1994; 56:23-8.
- (9) Ponta H, Hofmann M, Herrlich P. Recent advances in the genetics of metastasis. *Eur J Cancer* 1994;30A:1995-2001.
- (10) Degen WG, Weterman MA, van Groningen JJ, Cornelissen IM, Lemmers JP, Agterbos MA, et al. Expression of nma, a novel gene, inversely correlates with the metastatic potential of human melanoma cell lines and xenografts. *Int J Cancer* 1996;65:460-5.
- (11) Safarians S, Sternlicht MD, Freiman CJ, Huaman JA, Barsky SH. The primary tumor is the primary source of metastasis in a human melanoma/scid model. Implications for the direct autocrine and paracrine epigenetic regulation of the metastatic process. *Int J Cancer* 1996;66:151-8.
- (12) Welch DR, Chen P, Miele ME, McGary CT, Bower JM, Stanbridge EJ, et al. Microcell-mediated transfer of chromosome 6 into metastatic human C8161 melanoma cells suppresses metastasis but does not inhibit tumorigenicity. *Oncogene* 1994;9:255-62.
- (13) Miele ME, Robertson G, Lee JH, Coleman A, McGary CT, Fisher PB, et al. Metastasis suppressed, but tumorigenicity and local invasiveness unaffected, in the human melanoma cell line Me/JuSo after introduction of human chromosomes 1 or 6. *Mol Carcinog* 1996;15:284-99.
- (14) You J, Miele ME, Dong C, Welch DR. Suppression of human melanoma metastasis by introduction of chromosome 6 may be partially due to inhibition of motility, but not to inhibition of invasion. *Biochem Biophys Res Commun* 1995;208:476-84.
- (15) Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction [see comment citation in Medline]. *Science* 1992;257:967-71.
- (16) Liang P, Averboukh L, Pardee AB. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic Acids Res* 1993;21:3269-75.
- (17) van Groningen JJ, Bloemers HP, Swart GW. Identification of melanoma inhibitory activity and other differentially expressed messenger RNAs in human melanoma cell lines with different metastatic capacity by messenger RNA differential display. *Cancer Res* 1995;55:6237-43.
- (18) Lee SW, Tomasello C, Sager R. Positive selection of candidate tumor-suppressor genes by subtractive hybridization. *Proc Natl Acad Sci U S A* 1991;88:2825-9.
- (19) Duguid JR, Rohwer RG, Seed B. Isolation of cDNAs of scrapie-modulated RNAs by subtractive hybridization of a cDNA library. *Proc Natl Acad Sci U S A* 1988;85:5738-42.
- (20) Welch DR, Bisi JE, Miller BE, Conaway D, Seftor EA, Yohem KH, et al. Characterization of a highly invasive and spontaneously metastatic human malignant melanoma cell line. *Int J Cancer* 1991;47:227-37.
- (21) Trent JM, Stanbridge EJ, McBride HL, Meese EU, Casey G, Araujo DE, et al. Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science* 1990;247:568-71.
- (22) Gualandri F, Morelli C, Pavan JV, Rimessi P, Senzi A, Bonfatti A, et al. Induction of senescence and control of tumorigenicity in BK virus transformed mouse cells by human chromosome 6. *Genes Chromosomes Cancer* 1994;10:77-84.
- (23) Schweinfest CW, Henderson KW, Gu JR, Kottaridis SD, Besbas S, Panotopoulou E, et al. Subtraction hybridization cDNA libraries from colon carcinoma and hepatic cancer. *Genet Anal Tech Appl* 1990;7:64-70.
- (24) Welch DR, Neri A, Nicolson GL. Comparison of spontaneous and 'experimental' metastasis using rat 13762 mammary adenocarcinoma cell clones. *Invasion Metastasis* 1983;3:65-80.
- (25) Lee JH, Welch DR. Differential display and subtractive hybridization to identify highly expressed genes in metastasis-suppressed chromosome 6/human melanoma hybrids. *Proc Am Assoc Cancer Res* 1996; 37:530.
- (26) Cohen GB, Ren R, Baltimore D. Modular binding domains in signal transduction proteins. *Cell* 1995;80:237-48.
- (27) Alexandropoulos K, Cheng G, Baltimore D. Proline-rich sequences that bind to Src homology 3 domains with individual specificities. *Proc Natl Acad Sci U S A* 1995;92:3110-4.
- (28) von Heijne G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res* 1986;14:4683-90.
- (29) Hennessey C, Henry JA, May FE, Westley BR, Angus B, Lennard TW. Expression of the antimetastatic gene nm23 in human breast cancer: an association with good prognosis. *J Natl Cancer Inst* 1991;83:281-5.
- (30) Musacchio A, Wilmanns M, Saraste M. Structure and function of the SH3 domain. *Prog Biophys Mol Biol* 1994;61:283-97.
- (31) Joseph G, Pick E. Peptide walking is a novel method for mapping functional domains in proteins. Its application to the Rac1-dependent activation of NADPH oxidase. *J Biol Chem* 1995;270:29079-82.
- (32) Feng S, Chen JK, Yu H, Simon JA, Schreiber SL. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* 1994;266:1241-7.
- (33) Mayer BJ, Eck MJ. SH3 domains. Minding your p's and q's. *Curr Biol* 1995;5:364-7.
- (34) Rickles RJ, Botfield MC, Zhou XM, Henry PA, Brugge JS, Zoller MJ. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A* 1995;92:10909-13.
- (35) Ishino M, Ohba T, Sasaki H, Sasaki T. Molecular cloning of a cDNA encoding a phosphoprotein, Efs, which contains a Src homology 3 domain and associates with Fyn. *Oncogene* 1995;11:2331-8.
- (36) Koch CA, Anderson D, Moran MF, Ellis C, Pawson T. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* 1991;252:668-74.
- (37) Musacchio A, Noble M, Pauptit R, Wierenga R, Saraste M. Crystal structure of a Src-homology 3 (SH3) domain [see comment citation in Medline]. *Nature* 1992;359:851-5.
- (38) Pawson T, Gish GD. SH2 and SH3 domains: from structure to function. *Cell* 1992;71:359-62.
- (39) Pawson T, Schlessinger J. SH2 and SH3 domains. *Curr Biol* 1993;3:434-42.
- (40) Feller SM, Ren R, Hanafusa H, Baltimore D. SH2 and SH3 domains as molecular adhesives: the interactions of Crk and Abl. *TIBS Trends Biochem Sci* 1994;19:453-8.
- (41) Nobes CD, Hall A. Rho, Rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* 1995;81:53-62.
- (42) Ridley AJ. Membrane ruffling and signal transduction. *Bioessays* 1994; 16:321-7.
- (43) Cantley LC, Auger KR, Carpenter C, Duckworth B, Graziani A, Kapeller R, et al. Oncogenes and signal transduction [published erratum appears in *Cell* 1991;65:following 914]. *Cell* 1991;64:281-302.
- (44) Trent JM, Thompson FH, Meyskens FL Jr. Identification of a recurring translocation site involving chromosome 6 in human malignant melanoma. *Cancer Res* 1989;49:420-3.
- (45) Guan XY, Meltzer PS, Cao J, Trent JM. Rapid generation of region-specific genomic clones by chromosome microdissection: isolation of DNA from a region frequently deleted in malignant melanoma. *Genomics* 1992;14:680-4.
- (46) Millikin D, Meese E, Vogelstein B, Witkowski C, Trent JM. Loss of heterozygosity for loci on the long arm of chromosome 6 in human malignant melanoma. *Cancer Res* 1991;51:5449-53.

Notes

Present address: M. E. Miele, Department of Medical Technology, University of Delaware, Newark.

Supported by Public Health Service grants CA62168 and CA44470 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, by the National Foundation for Cancer Research, and by gifts from the Jake Gittlen Memorial Golf Tournament.

This paper is dedicated to the memory of Ray C. Welch, who was always supportive and encouraging.

Manuscript received May 10, 1996; revised August 14, 1996; accepted September 10, 1996.